

**УПРАВЛЕНИЕ  
ИНФОРМАЦИОННЫМИ РИСКАМИ  
И ОБЕСПЕЧЕНИЕ БЕЗОПАСНОСТИ  
ИНФОКОММУНИКАЦИОННЫХ СИСТЕМ**

Сборник научных трудов

**Под редакцией член-корреспондента РАН  
В.И. Борисова**

**Выпуск 1 (11), 2016**

**Воронеж  
Издательство «Научная книга»  
2016**

**УДК 004.42**

**ББК**

**У**

Управление информационными рисками и обеспечение безопасности инфокоммуникационных систем: Сб. науч. тр.; под ред. чл.-корр. РАН В.И. Борисова. Вып. 1 (11), 2016. – Воронеж: Издательство «Научная книга». 2016. – 97 с.

**ISBN**

Статьи посвящены широкому кругу вопросов, связанных с исследованиями в области обеспечения безопасности процессов, протекающих в инфокоммуникационных системах, посредством оценки и регулирования их рисков.

Издание предназначено для специалистов в области обеспечения безопасности информационных систем и процессов различных классов.

**Редколлегия:** А.Г. Остапенко (ответственный редактор),  
Е.В. Субботина (ответственный секретарь),  
А.В. Бабурин, В.И. Белоножкин, В.Н. Деревянко, С.А. Ермаков,  
В.Л. Каркоцкий, А.О. Калашников, Д.О. Карпеев, А.Н. Мокроусов,  
Е.А. Москалева, И.П. Нестеровский, Г.А. Остапенко, О.А. Остапенко,  
А.С. Пахомова, О.В. Поздышева, Н.М. Радько, К.А. Разинкин,  
Д.Н. Рахманин, А.Ю. Савинков, Н.М. Тихомиров, Н.Н. Толстых,  
О.Н. Чопоров, Е.Ю. Чапурин, В.Б. Щербаков, Ю.К. Язов.

Статьи сборника издаются в авторской редакции.

**УДК 004.42**

**ББК**

**У**

©Коллектив авторов, 2016

## ОГЛАВЛЕНИЕ

СТРУКТУРНО – ФУНКЦИОНАЛЬНАЯ СПЕЦИФИКА РАСПРОСТРАНЕНИЯ КОНТЕНТА В СОЦИАЛЬНОЙ СЕТИ ДЛЯ ОБЩЕНИЯ FACEBOOK К.В. Зайцев, А.В. Паринов .....	4
ИССЛЕДОВАНИЕ СТРУКТУРНО – ФУНКЦИОНАЛЬНОЙ СХЕМЫ СОЦИАЛЬНОЙ СЕТИ ДЛЯ ОБЩЕНИЯ ВКОНТАКТЕ Т.И. Дубинская, А.В. Паринов.....	25
СТРУКТУРНО – ФУНКЦИОНАЛЬНАЯ СПЕЦИФИКА РАСПРОСТРАНЕНИЯ КОНТЕНТА В СОЦИАЛЬНОЙ FOURSQUARE С.И. Уколов, А.В. Паринов .....	44
СТРУКТУРНО – ФУНКЦИОНАЛЬНАЯ СПЕЦИФИКА РАСПРОСТРАНЕНИЯ КОНТЕНТА В СОЦИАЛЬНОЙ СЕТИ TOUROUT Д.И. Филиппов, А.В. Паринов .....	63
НАУЧНО-МЕТОДИЧЕСКОЕ ОБОБЩЕНИЕ ПРОЦЕДУР МОДЕЛИРОВАНИЯ ИНФОРМАЦИОННОЙ ДИФФУЗИИ В СОЦИАЛЬНЫХ СЕТЯХ А.В. Паринов, В.В. Морковина, А.А. Остапенко.....	86

**УДК 004.056.57**

**СТРУКТУРНО – ФУНКЦИОНАЛЬНАЯ СПЕЦИФИКА  
РАСПРОСТРАНЕНИЯ КОНТЕНТА В СОЦИАЛЬНОЙ  
СЕТИ ДЛЯ ОБЩЕНИЯ FACEBOOK**

**К.В. Зайцев, А.В. Паринов**

Колоссальная величина активности социальных сетей является наиболее привлекательным фактором для распространения вредоносного контента. Facebook не является исключением. Будучи одной из наиболее предпочтительных социальных сетей для просмотра новостей, Facebook также является и одной из самых прибыльных для злоумышленников [4,11]. Наиболее излюбленный способ последних – распространение фишинговых ссылок во время крупных мировых событий [4].

Проблема идентификации вредоносного контента не является специфической особенностью Facebook и была широко изучена во многих социальных сетях. Исследователи разрабатывали и изучали модели для обнаружения спама и других видов опасного медиаконтента [17], но многие из подходов, используемых для других социальных сетей, не могут быть непосредственно перенесены на Facebook, так как они в значительной степени зависят от данных, которые не являются публично доступными из Facebook [10]. К ним относятся профиль, сетевая информация, возраст аккаунта, общее количество сообщений, количество социальных связей и т.д.

Таким образом, полный анализ структуры сети, выделение субъектов и объектов взаимодействующих в ней с последующим выводом алгоритма для управления рисками и прогнозирования развития вирусных эпидемий представляет наиболее актуальный вектор исследования социальных сетей в настоящий момент.

Рассмотрим общую классификацию, охватывающую большую часть информации, потенциально возможной к использованию злоумышленниками в сети Facebook [10].

Facebook можно разбить по трем критериям:

1. По содержанию.
2. По форме подачи информации.
3. По критерию безопасности просмотра и использования.

Рассмотрим подробнее данную классификацию (рис. 1).



Рис. 1. Классификация циркулирующего контента

Вредоносный контент, в зависимости от вида, может оказывать негативное воздействие на пользователя и манипулировать данными его устройств, используемых для взаимодействия в сети. Рассмотрим виды вредоносного контента сети Facebook [14,19]:

1. Вредоносное ПО. Согласно исследованию, 36 процентов пользователей Facebook сообщили о получении вредоносного контента через социальную сеть. Киберпреступники используют различные методы

для распространения вредоносного программного обеспечения в социальных сетях. Вредоносные посты, личные сообщения или ссылки, размещенные в профилях Facebook являются общими способами средствами передачи вредоносных программ другим пользователям.

2. Спам. При подготовке к спам-атаке злоумышленники регистрируют поддельные учетные записи на сайтах социальных сетей для последующей рассылки сообщений. Тем не менее, в настоящее время они нашли способ для атаки на целевую жертву. Так как вероятность, что типичный пользователь откроет сообщение от неизвестного отправителя мала, они крадут чужие аккаунты и отправляют спам-сообщения от их имени всему контактному списку жертвы, увеличивая вероятность установления диалога.

3. Фишинг. Также как спам и вредоносное ПО, фишинг опирается на доверительные отношения между пользователями сети. Типичная фишинговая операция заключается в отправлении сообщений, содержащих ссылку на фальсифицированную страницу регистрации Facebook, которая используется злоумышленниками для получения данных аккаунта жертвы.

4. Мошенничество. Представляет собой выманивание информации о реквизитах доступа к системе (в отличие от взлома самой системы). В социальных сетях люди чувствуют себя более комфортно и охотнее делятся персональной информацией. Это повышает риск того, что злоумышленник раскроет конфиденциальные данные через социальную сеть.

5. Ботнеты. Ботнет создается путем компрометации некоторого количества компьютеров, соединенных с Интернетом. Атакующие могут контролировать ботнет с помощью сообщений, отправляемых через социальные сети. Был проведен эксперимент по превращению Facebook в ботнет: специальное приложение передавало http-запросы на скомпрометированный хост, и оказалось, что если бы атакующий разослал

вирус нескольким миллионам пользователей, то атакованный хост ежедневно раздавал бы по Сети сотни гигабайт вредоносных файлов.

Будучи одной из самых крупных социальных сетей Facebook обладает высоким уровнем защищённости [14]. Опорным уровнем защиты является использование технологии Secure Socket Layer (SSL), шифрующей вводимые пользователями данные на клиентской стороне и расшифровывающей их на серверной. Таким образом, наиболее распространённую атаку «Man in the middle», представляющую собой прослушивание канала передачи данных между корреспондентами, невозможно осуществить. Серверы, на которых хранятся пользовательские данные, имеют встроенные системы защиты, включающие в себя новейшее антивирусное ПО, системы обнаружения вторжений, брандмауэры, а также мощные системы резервации. Данные методы защиты противодействуют внедрению киберпреступниками распространённого вредоносного ПО.

Рассмотренный контент может содержаться в различных ресурсах сети.

Для дальнейшего анализа сети, их необходимо классифицировать (рис.2):

1. Ресурсы коллективного использования с настройками приватности. К ним относятся ресурсы с возможностью настраивать список пользователей, имеющих к ним доступ. К ним относятся: группа, стена и профиль.

2. Ресурсы коллективного использования. Все пользователи социальной сети имеют к ним доступ. Они могут быть представлены страницами, сообществами и статьями.

3. Ресурсы персонального использования. Представляют данные пользователя, доступные только с его учётной записи. Возможность открыть доступ к данным ресурсам другим пользователям сети

отсутствует. К данному виду ресурсов относятся: личные данные, новостная лента, закладки и документы.

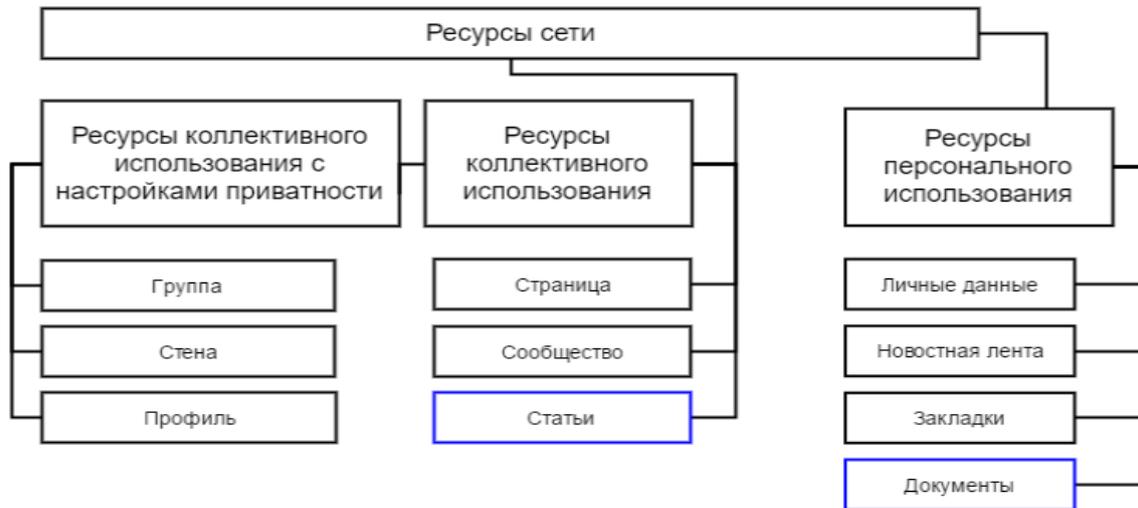


Рис. 2. Сетевые ресурсы для размещения контента

Рассмотрим возможные действия субъектов сети при взаимодействии с объектами. Их можно разделить по трем критериям: размещение, ознакомление и реагирование (рис. 3).



Рис. 3. Классификация действий субъектов сети Facebook

Рассмотрим подробнее процесс обмена контентом между сетями. Для этого все интернет-ресурсы разделим на две части. Первая представляет собой отслеживаемую область, представляемую рассматриваемой нами социальной сетью и неотслеживаемую область, агрегирующую остальные ресурсы. Любой авторизированный пользователь имеет возможность в один клик поделиться информацией с внешнего источника со своими друзьями или подписчиками. Это достигается благодаря высокому уровню интеграции Facebook с другими популярными сетевыми платформами. Также возможен и обратный процесс, позволяющий поделиться информацией с Facebook в другой социальной сети. Рассмотрим схему, отражающую данные процессы (рис. 4).



Рис. 4. Движение контента между сетями

Рассмотрим связь между субъектами, функционирующими в сети, и возможными действиями по управлению контентом. Все субъекты социальной сети Facebook имеют базовый функционал, который может быть расширен. Например, неавторизованный пользователь получает воз-

возможность не только просматривать контент, но и делиться своим контентом после прохождения процесса авторизации. Рассмотрим роли субъектов и их возможности по управлению контентом (рис. 5).

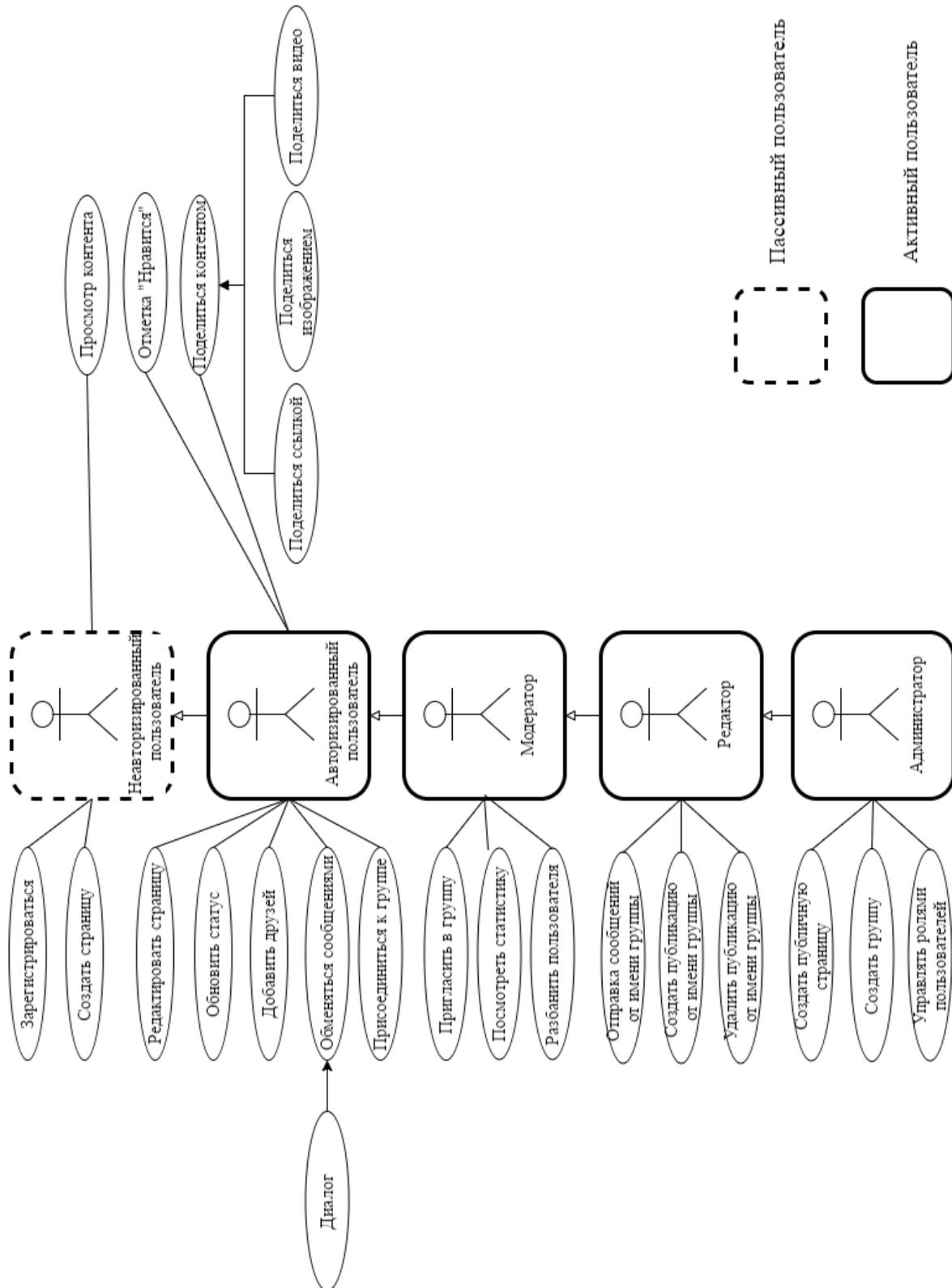


Рис. 5. Роли пользователей в Facebook

Рассмотрим подробнее взаимодействие компонентов сети. На вершине иерархии ролей пользователей располагается Администрация сети. Она имеет возможность управлять ролями пользователей, изменяя их права доступа к любому виду контента, а также устанавливать настройки доступа по размещению контента как из Facebook в другие социальные сети, так и в обратном направлении. У пользователя, получившего роль Модератора, появляется возможность управлять группами, в частности:

1. Закрывать возможности по вступлению в группу без приглашения со стороны её участников.
2. Удалять чужие посты со стены группы.
3. Приглашать пользователей в закрытую группу.
4. Исключать пользователей.

Основной задачей Редактора является добавление постов в группу, закрытую настройками приватности со стороны Администрации, а также публикация постов в группе по запросам пользователей. Возможности активных пользователей:

1. Редактирование, удаление, создание постов на собственной странице, в группе, в новостной ленте.
2. Дублирование контента со страниц других пользователей, групп или других социальных сетей (функция «поделиться»).
3. Редактирование собственного профиля.
4. Сохранение текста, анимаций в собственные документы.
5. Добавление групп, профилей пользователей, постов в закладки

Возможности неавторизованных пользователей: регистрация в социальной сети и авторизация.

Основываясь на полученных результатах, построим структурно-функциональную схему, отражающую специфику взаимодействия пользователей в сети Facebook (рис. 6):

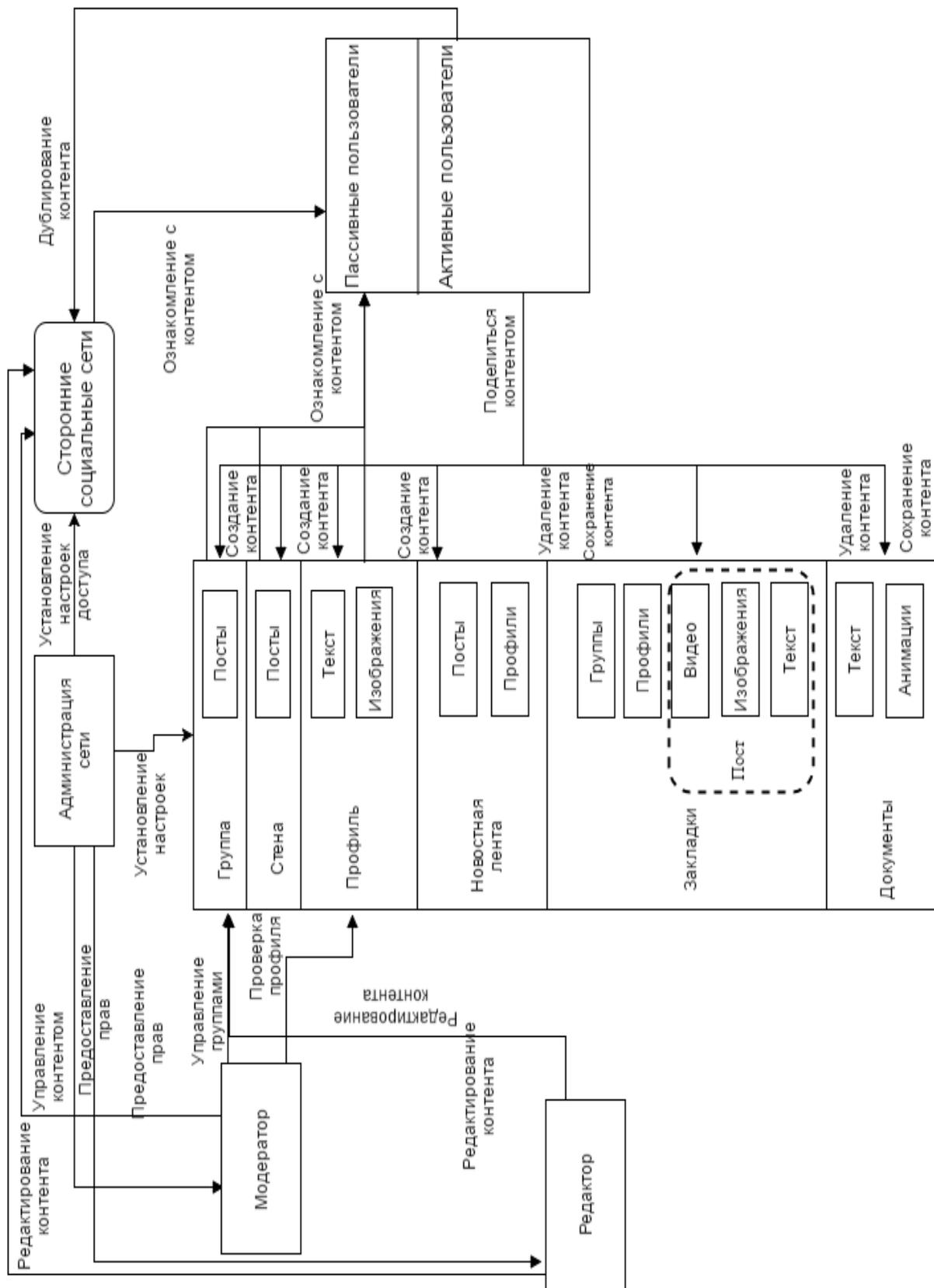


Рис. 6. Структурно-функциональная схема взаимодействия компонентов сети.

Статистические данные, представленные сайтом konect [16] для сети Facebook представляют собой трехместные предикаты, на первом месте которых расположена вершина, из которой исходит ребро, на втором – вершина в которую оно заходит, а на третьем количество трафика определённой ценности в единицу времени. Вершины представляют собой пользователей сети, а трафик – количество репостов, осуществлённых первым авторизованным пользователем со страницы второго авторизованного пользователя за 1 месяц [16].

Рассмотрим характеристику статистических данных (табл. 1):

Таблица 1 – Статистические данные социальной сети Facebook

Характеристические данные	Показатели
Количество узлов	482128
Количество ребер	7015397
Диаметр	34
Средняя длина пути	7,62
Коэффициент кластеризации (%)	31,28

На рисунке ниже представлен граф, полученный на основе исходных данных (рис. 7).

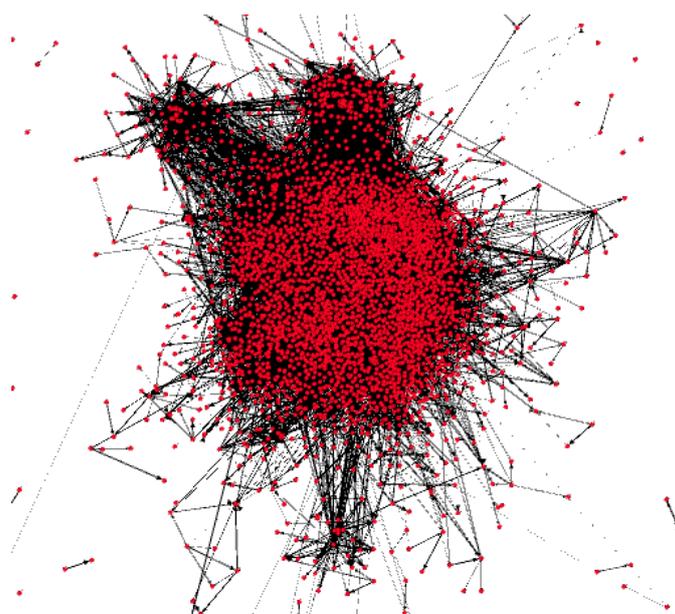


Рис. 7. Визуализация исходного графа

На данном графе явно прослеживается кластерность. Следствием является высокий коэффициент кластеризации. Это связано с тем, что пользователи сети Facebook могут ограничивать видимость своих публикаций или групп, в которых они являются администраторами, настройками приватности.

Трёхместные предикаты, представляющие статистические данные по социальной сети Facebook сайта kopect [16] описываются формулой (1):

$$\Gamma(x_i, a_{ij}, x_j) \Leftrightarrow \Gamma(i, \delta(a_{ij}), j), \quad (1)$$

где  $i$  и  $j$  – номера вершин  $x_i$  и  $x_j$  в сети,  $\delta(a_{ij})$  - вес дуги  $a_{ij}$ , связывающей  $x_i$  и  $x_j$ , и направленной от  $i$  к  $j$ .

Величина веса дуги представляет собой количество постов размещённых со страницы одного авторизованного пользователя на страницу другого при помощи функции репост в течение 1 месяца [13,15]. Данная величина рассчитывается по формуле (2):

$$\delta(a_{ij}) = K_{a_{ij}}, \quad (2)$$

где  $K_{a_{ij}}$  – количество репостов.

Такой формат позволяет построить звездную матрицу, элементы строки которой соответствуют дугам, входящим в данную вершину, а элементы столбца – дугам, исходящим из вершины.

Построим звездную матрицу взвешенной связности исходной сети (табл. 2):

Таблица 2 – Звездная матрица взвешенной связности исходной сети

Номер узла	1	2	3	4	...	1856	1857	...	482128
1	0	18	20	6	...	0	0	...	0
2	2	0	12	0	...	0	0	...	0
3	9	4	0	0	...	0	0	...	0
4	0	2	0	0	...	16	0	...	0

Продолжение таблицы 2

...	...	...	...	...	...	...	...	...	...
8	5	3	0	1	...	0	0	...	0
9	17	0	0	2	...	0	0	...	0
...	...	...	...	...	...	...	...	...	...
482128	0	0	0	0	...	0	0	...	0

Для последующей репрезентативной выборки из мультиразмерной сети необходимо знать удельный вес ее вершин и дуг. Для этого пронормируем их веса по сумме весов всех дуг сети по формуле (2):

$$\sum_{\substack{i,j \\ i \neq j}} \delta(a_{ij}). \quad (2)$$

Пронормируем полученную величину (3):

$$\delta(\bar{a}_{ij}) = \delta(a_{ij}) / \sum_{\substack{i,j \\ i \neq j}} \delta(a_{ij}), \quad (3)$$

которая показывает удельный вес трафика в дуге  $a_{ij}$  к суммарному трафику сети. Она и будет характеризовать степень ее взвешенной (по трафику) центральности.

Суммарный трафик сети не разделяет входящие и исходящие дуги. Поэтому для определения взвешенной центральности вершины  $x_s$  используем сумму (4):

$$\sum_i \delta(a_{si}) + \sum_j \delta(a_{js}). \quad (4)$$

Пронормируем полученную сумму по суммарному трафику сети используя формулу (5):

$$\delta(\bar{x}_s) = [\sum_i \delta(a_{si}) + \sum_j \delta(a_{js})] / \sum_{\substack{i,j \\ i \neq j}} \delta(a_{ij}), \quad (5)$$

которая будет характеризовать удельный вес трафика, проходящего через вершину  $x_s$ , по отношению ко всему трафику сети. Это и будет характеризовать степень взвешенной (по трафику) центральности данной вершины. Исходя из вышеизложенного, можно построить квадратную

матрицу взвешенной центральности элементов исходной сети, играющую ключевую роль в формировании репрезентативной выборки (табл. 3):

Таблица 3- Квадратная матрица взвешенной центральности элементов сети

Номер узла	1	2	3	...	278	279	...	482128
1	0.00044	0.000034	0.000037	...	0	0	...	0
2	0	0.000167	0	...	0	0	...	0
3	0.00008	0	0.00019	...	0	0	...	0
...	...	...	...	...	0	0	...	...
278	0	0	0	...	0.000153	0	...	0
279	0	0	0	...	0	0.00061	...	0
...	...	...	...	...	...	...	...	...
482128	0	0	0	...	0	0	...	0.00000003

Высокая сложность проведения анализа социальной сети Facebook объясняется огромным количеством вершин и связей между ними в используемых статистических данных [11] и вызывает необходимость в получении выборки, сохраняющей свойства сети с заданной точностью.

Для получения репрезентативной выборки воспользуемся алгоритмом, реализованным совместно с партнерами по комплексной работе. За допустимую погрешность примем не более 5% трафика для усеченной модели сети.

Алгоритм получения репрезентативной выборки состоит из четырёх шагов:

1. Вычленим из исходной матрицы взвешенной центральности элементов сети ее диагонали. Оставшаяся усеченная матрица с обнулёнными диагональными элементами будет использована на шаге 4.

2. Отсортируем полученную матрицу взвешенной центральности по убыванию значений степени взвешенной центральности вершин  $\delta(\bar{x}_s)$ , расположенных на главной диагонали.

3. Двигаясь от конца к началу значений главной диагонали, найдем минимальную сумму значений  $\delta(\bar{x}_s)$ , такую что  $\sum \delta(\bar{x}_s) \approx 0.05$ , так как допустимая потеря составляет 5% трафика.

4. Удалим из усеченной матрицы все элементы, входящие в сумму, рассчитанную на предыдущем шаге путём обнуления столбцов и строк.

Применив данный алгоритм к построенной матрице взвешенной центральности построим проранжированную усеченную матрицу социальной сети Facebook (табл. 4):

Таблица 4 – Проранжированная усеченная матрица сети Facebook

Но- мер узла	157	46	597	...	332	...	3644	2267	1374
157	0	0	0	...	0	...	0	0	0
46	0	0	0,00017	...	0	...	0	0	0
597	0	0,0003	0	...	0	...	0,0004	0	0
...	...	...	...	...	...	...	...	...	...
1302	0	0	0	...	0	0	0	...	0
836	0,00031	0	0	...	0	0	0	...	0
332	0	0	0,00327	...	0	0	0	0	0,00015
...	...	...	...	...	...	...	...	...	...
3644	0	0	0	...	0	...	0	0	
2267	0	0	0	...	0,00018	...	0	0	0
1374	0	0,0012	0	...	0	...	0	0,000435	0

Таким образом, полученная матрица будет представлять собой вы-  
борку, в которой отражена связность всех ее компонентов. С помощью  
программного обеспечения Gephi [21], визуализируем полученный резуль-  
тат (рис 8.).

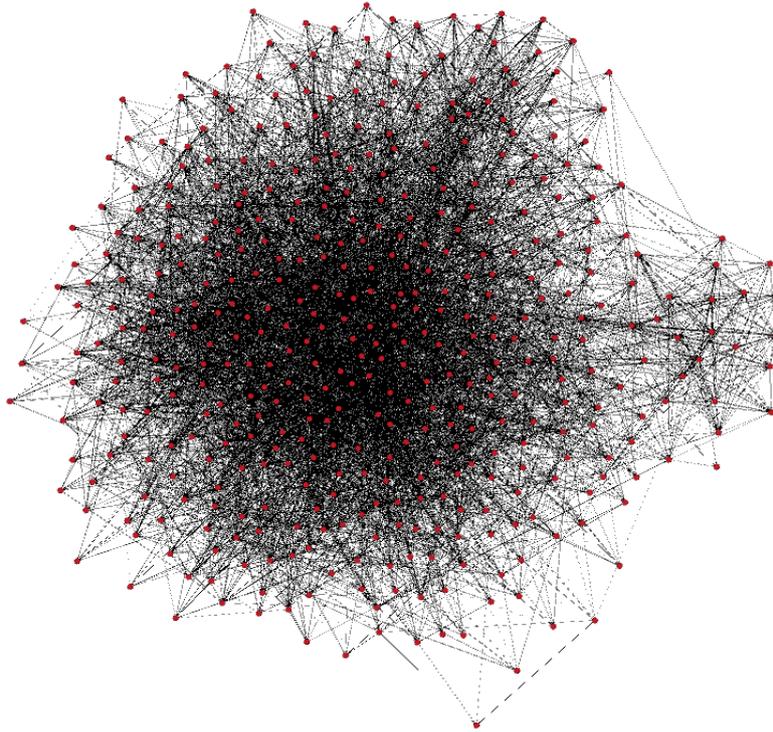


Рис. 8. Визуализация графа на основе полученной выборки

Обратим внимание, что кластерность, присущая исходному графу, сохранилась в выборке, что подтверждает верность расчётов по выбранному алгоритму.

Осуществим с помощью программного обеспечения Gephi [21] оценку параметров графа выборки (табл. 5):

Таблица 5 - Параметры графа выборки

Характеристические данные	Показатели
Количество узлов	410
Количество ребер	17520
Диаметр	24
Средняя длина пути	3,23
Коэффициент кластеризации (%)	25,76

Сформируем модифицированную матрицу смежности социальной сети Facebook являющейся базовой при моделировании процессов информационной диффузии в сети Facebook (табл. 5):

Таблица 6 - Модифицированная матрица смежности сети Facebook

Номер узла	157	46	597	...	332	...	3644	2267	1374
157	43	0	0	...	0	...	0	0	0
46	0	53	1	...	0	...	0	0	0
597	0	1	31	...	0	...	1	0	0
...	...	...	...	...	...	...	...	...	...
332	0	0	-1	...	21	...	0	0	1
...	...	...	...	...	...	...	...	...	...
3644	0	0	0	...	0	...	5	0	0
2267	0	0	0	...	-1	...	0	2	0
1374	0	-1	0	...	0	...	0	-1	4

Данная матрица строится на основе двух правил:

1. Недиagonальные элементы равны  $\pm 1$ , так как интересует лишь факт смежности вершин истоков (+1) и стоков (-1).
2. Диагональные элементы имеют вес  $k$ , равный сумме элементов (+1) столбца.

Таким образом, необходимые матрицы для последующего анализа сети были построены.

Построим таблицу распределения генеральной совокупности, представляющую варианты данной совокупности (степени вершин) и их частоту (табл. 6).

Таблица 7 – Генеральная совокупность

$i$	1	2	3	..	48	49	50	...	124	125	126
$x_i$	1	2	3	..	50	51	53	...	158	169	188
$n_i$	183524	54395	31562	..	430	602	516	...	43	41	34

Рассмотрим выборочную совокупность, описывающую вершины, полученные оставшиеся после применения алгоритма выборки (табл. 7):

Таблица 8 – Выборочная совокупность

$i$	1	2	3	...	18	19	20	...	39	40	41
$x_i$	1	2	3	...	50	51	53	...	81	85	128
$n_i$	8	10	14	...	14	18	22	...	2	1	1

Для проведения дальнейших эпидемиологических исследований необходимо доказать подобие полученной выборки исходной генеральной совокупности. Доказательство репрезентативности осуществляется в 4 шага:

1. Визуализация выборки и оценка сохранения кластерности исходного графа (рис. 8).
2. Построение графика зависимостей степеней вершин от долей вершин для генеральной и выборочной совокупностей для предварительной оценки сохранения исходного распределения.
3. Доказательство принадлежности выборочной совокупности степенному распределению с использованием критерия Пирсона.
4. Расчёт среднеквадратического отклонения генеральной совокупности от полученной выборочной совокупности.

На основе полученных генеральной и выборочной совокупностей проведём сравнение распределений количества вершин степеней для исходного графа и для графа, представляемого выборкой (рис. 9).



Рис. 9. График зависимости степеней вершин от их количества

На представленном выше графике прослеживается сохранение закона распределения степеней вершин в выборочной совокупности.

Воспользуемся критерием Пирсона для доказательства принадлежности выборки степенному распределению. Для этого рассчитаем  $\chi^2_{\text{набл}}$ , используя формулу (6):

$$\chi^2_{\text{набл}} = \sum_{i=1}^k \frac{(n - np_i)^2}{np_i}, \quad (6)$$

где  $n$  – объем выборки,  $p_i$  - вероятность попадания случайной величины  $X$ .

Для выборочной совокупности  $\chi^2 = 41.13$ , что соответствует объёму выборки  $n = 41$  и числу степеней свободы  $\nu = 40$ . Зададим уровень значимости  $\gamma = 0,05$ , тогда  $p = 0,95$ . Табличное значение  $\chi^2(41)$  равно 50.99338. Таким образом, принадлежность выборочной совокупности степенному распределению доказана, так как табличное значение  $\chi^2$  больше вычисленного [22].

В соответствии с алгоритмом доказательства репрезентативности выборки, рассчитаем среднеквадратическое отклонение генеральной от выборочной совокупности по формуле (7):

$$\sigma = \sqrt{\frac{1}{n} \sum_{i=1}^n (X_i - Y_i)^2}, \quad (7)$$

где:  $n$  – объем выборки,  $X_i$  –  $i$ -ый элемент генеральной совокупности,  $Y_i$  –  $i$ -ый элемент выборочной совокупности.

В результате вычисления по формуле (10) на основе построенных таблиц распределений совокупностей, получим:  $\sigma = 0,0729$ .

Таким образом, полученная выборка является репрезентативной с точностью приближённо равной 93%, что позволяет использовать её для

дальнейших эпидемиологических исследований [2-8], а также для оценки и регулирования рисков [4], возникающих в социальной сети Facebook.

#### Литература:

1. Analytical estimation of the component viability of distribution automated information data system / G.A. Ostapenko, D.G. Plotnicov, O.Y. Makarov, N.M. Tikhomirov, V.G. Yurasov // World Applied Sciences Journal. – 2013. – 25 (3). – P. 416-420.
2. Analytical models of information-psychological impact of social information networks on users / G.A. Ostapenko, L.V. Parinova, V.I. Belonozhkin, I.L. Bataronov, K.V. Simonov // World Applied Sciences Journal. – 2013. – 25 (3). – P. 410-415.
3. Assessment of the system's EPI-resistance under conditions of information epidemic expansion / N.M. Radko, A.G. Ostapenko, S.V. Mashin, O.A. Ostapenko, D.V. Gusev // Biosciences Biotechnology Research Asia. – 2014. – Vol. 11 (3). – P. 1781-1784.
4. Benevenuto F. Characterizing user behavior in online social networks / F. Benevenuto, T. Rodrigues, M. Cha, and V. Almeida / In Proc. of ACM SIGCOMM Internet Measurement Conference. ACM, 2009. – P. 28-32.
5. Caldarelli G. Scale-Free Networks. Complex Webs in Nature and Technology / G. Caldarelli / Cambridge University Press, 2007. – P. 45-51.
6. Discreet risk-models of the process of the development of virus epidemics in non-uniform networks / V.V. Islamgulova, A.G. Ostapenko, N.M. Radko, R.K. Babadzhanov, O.A. Ostapenko // Journal of Theoretical and Applied Information Technology. – 2016. – P. 306-315.

7. Evolution of Networks: From Biological Networks to the Internet and WWW / S.N. Dorogovtsev, J.F.F. Mendes // Oxford University Press. – 2003. – P. 145 - 280.
8. Flood-attacks within the hypertext information transfer protocol: damage assessment and management / A.G. Ostapenko, M.V. Bursa, G.A. Ostapenko, D.O. Butrik // Biosciences Biotechnology Research Asia. – 2014. – Vol. 11 (Spl.End). – P. 173-176.
9. Higher order clustering effective coefficients in Barabasi - Albert networks / A. Fronczak, J.A. Holyst, M. Jedynek, J. Sienkiewicz // Physica A. – 2002. – P. 688- 694.
10. Understanding the user generated content and interactions on a Facebook brand page / Irena Pletikosa Cvijikj // ETH Zurich. – 2011. – P. 5-11.
11. Online Social Networks: Human Cognitive Constraints in Facebook and Twitter Personal Graphs / Waltham // Elsevier Inc. – 2015. – P. 1-14.
12. Optimization of expert methods used to analyze information security risk in modern wireless networks / S.A. Ermakov, A.S. Zavorykin, N.S. Kolenbet, A.G. Ostapenko, A.O Kalashnikov // Life Science Journal. – 2014. – № 11(10s). – P. 511-514.
13. Peak risk assessing the process of information epidemics expansion / N.M. Radko, A.G. Ostapenko, S.V. Mashin, O.A. Ostapenko, A.S. Avdeev // Biosciences Biotechnology Research Asia. – 2014. – Vol. 11. – P. 251-255.
14. Plannable Detecting Malicious Content on Facebook / Cybersecurity Education and Research Centre Prateek Dewan, Ponnurangam Kumaraguru // Malicious Content Research. – 2015. – Vol. 1. – P. 12-25.
15. Are all Social Networks Structurally Similar. / Sallaberry A. // A Comparative Study. – 2014. - P. 25-36.
16. База данных статистики по социальным сетям. - Электрон. Дан. – Режим доступа: <http://konect.uni-koblenz.de/>

17. Исследование вредоносного контента в социальных сетях. - Электрон. Дан. – Режим доступа: <https://www.sophos.com/en-us/press-office/press-releases/2010/02/security-report-2010.aspx>
18. Об устойчивости и мощности критериев проверки однородности средних / Лемешко Б.Ю., Лемешко С.Б. // Измерительная техника. – 2008. №9. – С. 2-4.
19. Распространение нежелательной информации в социальных сетях Интернета /Абрамов К. Г., Абрамов, Ю.М. // Диссертация – 2010. – С. 45-48.
20. Социальные сети: модели информационного влияния, управления и противоборства / Губанов Д.А., Новиков Д.А., Чхартишвили А.Г. // Учебное издание. – С. 203-205.
21. Средство визуализации данных. - Электрон. Дан. – Режим доступа: <https://gephi.org/>
22. Теория вероятностей и математическая статистика / Гмурман В.Е. // Учебное пособие. – 2007. – С. 142-151.

Воронежский научно-образовательный центр  
управления информационными рисками

**УДК 004.056.57**

## **ИССЛЕДОВАНИЕ СТРУКТУРНО – ФУНКЦИОНАЛЬНОЙ СХЕМЫ СОЦИАЛЬНОЙ СЕТИ ДЛЯ ОБЩЕНИЯ ВКОНТАКТЕ**

**Т.И. Дубинская, А.В. Паринов**

Социальная сеть — это социальная структура, состоящая из узлов, связанных между собой одним или несколькими способами посредством социальных взаимоотношений (примерами узлов могут быть отдельные люди, группы людей или сообщества). В обычном смысле этого слова социальная сеть — это сообщество людей, связанных общими интересами, общим делом или какими-то другими причинами для непосредственного общения [8,9].

Социальная сеть ВКонтакте представляет собой ресурс с огромнейшей популярностью, большим объемом всевозможного контента и развитым функционалом. Основным назначением сети является функционирование в качестве механизма коммуникации людей между собой, обеспечение возможности поиска нужной персоны или сообщества [9].

ВКонтакте – социальная сеть, занимающая 3 место по популярности в России. На данный момент количество посещений данной социальной сети в месяц превышает 48 миллионов человек и около 22 миллионов человек в сутки [8]. ВКонтакте является очень удобным средством для общения, распространения информации, поиска людей, групп по интересам, компаний, и даже ведения интернет-бизнеса.

Пользователи социальной сети ВКонтакте могут размещать создавать и публиковать контент в различных разделах. Среди разделов для размещения контента выделяются три основные группы:

- Ресурсы коллективного использования с настройками приватности.
- Ресурсы коллективного использования.
- Ресурсы персонального использования.

На рисунке 1 представлена схема сетевых ресурсов для размещения контента.

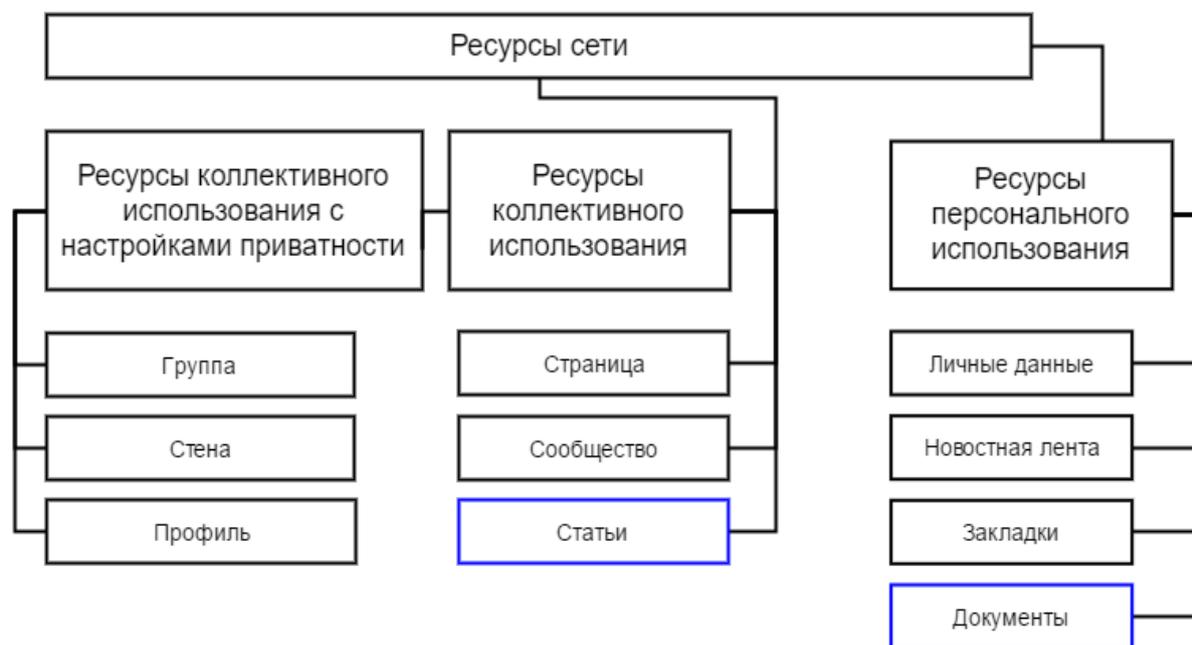


Рис. 1. Сетевые ресурсы для размещения контента

К ресурсам коллективного использования с настройками приватности относятся группа, стена и профиль. Группы могут быть как открытыми, так и закрытыми. В открытых группах пользователям, не являющимся участниками группы, может быть доступна не вся информация. Если же пользователь является участником, то ему доступен больший перечень возможностей. Что касается закрытых групп, то для просмотра, размещаемого в ней контента, пользователь обязательно должен быть участником группы. Стена представляет собой записи, которые расположены в хронологическом порядке на странице пользователя. Пользователь может ограничивать доступ к стене: запретить добавление записей, просмотр записей,

которые были добавлены другими пользователями, ограничить возможность комментирования или же совсем закрыть стену [8,9]. Так же пользователь может ограничивать доступ к своему профилю, может выбрать у кого будет возможность писать ему личные сообщения, кто сможет приглашать его в различные сообщества и приложения, кто может добавлять его в друзья [9].

К ресурсам персонального пользования относятся личные данные, новостная лента, закладки и документы. Весь контент, который находится в данных ресурсах доступен исключительно пользователю [8,9]. Он может просматривать изменения и обновления сообществ и его друзей в новостной ленте, а понравившийся контент добавлять в закладки.

Страница, статьи и сообщества входят в состав ресурсов коллективного пользования.

Особенностью социальной сети для общения ВКонтакте являются такие ресурсы, как статьи и документы. В статьях пользователь может ознакомиться с публикациям новостных изданий, политиков, знаменитостей [8,9]. В них собрана самая актуальная и обсуждаемая информация. В документах хранятся файлы, которые пользователь когда-либо загружал на сайт. Они могут быть представлены в различных форматах: текстовые, изображения, видео, анимации, архивы и другие.

Что касается пользователей в данной социальной сети, то их можно разделить на пять основных групп:

- 1) Администраторы. Отвечают за управление сообществом.
- 2) Модераторы. Следят за порядком.
- 3) Редакторы. Публикуют контент.
- 4) Авторизованные пользователи. В данный перечень входят пользователи социальной сети, прошедшие процедуру регистрации аккаунта.

- Активные пользователи.
- Пассивные пользователи.

5) Неавторизованные пользователи. Оставшаяся масса людей, которые имеют ограниченные возможности в данной социальной сети. Им доступна возможность просмотра контента, неограниченного настройками приватности, а также предоставляется возможность зарегистрироваться в данной социальной сети [8,9].



Рис. 2. Субъекты, функционирующие в сети

Контент, циркулирующий в ресурсах социальной сети так же можно разбить на группы: по содержанию, по представлению и по критерию безопасности просмотра и использования.

На рисунке представлена более подробная классификация контента.

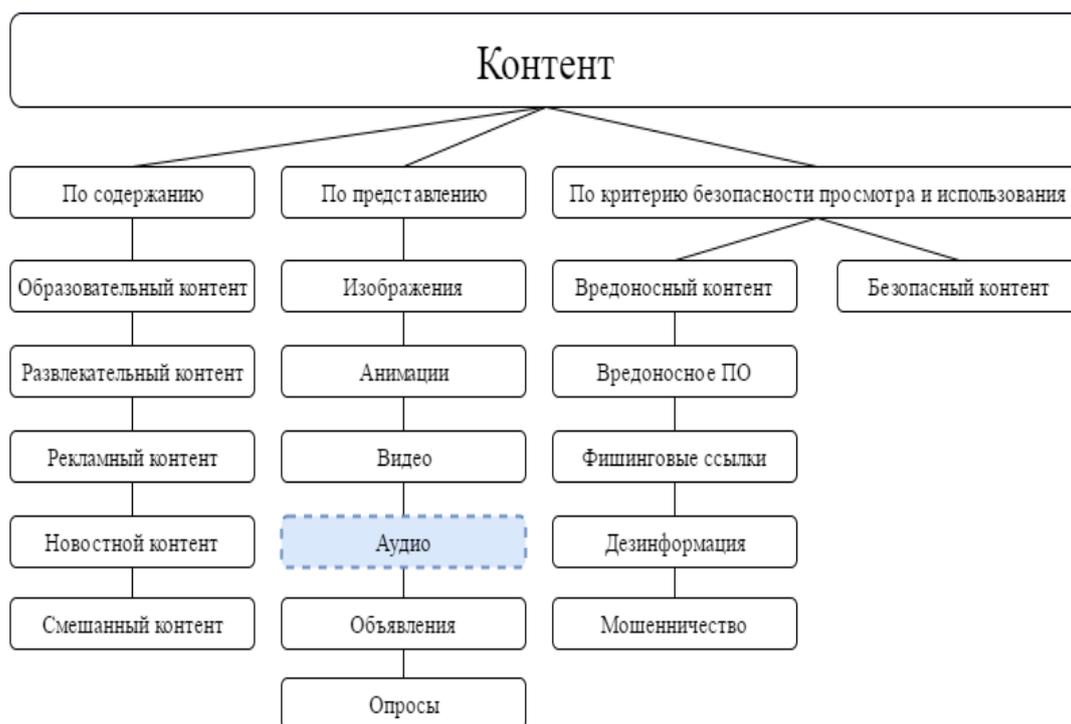


Рис. 3. Классификация контента в сети

**Образовательный контент.** К данному разделу можно отнести посты, содержащие обучающую информацию: книги, статьи, фото и видео уроки, мастер-классы, ссылки на различные образовательные ресурсы [8,9].

**Развлекательный контент.** Посты, содержащие в себе этот вид контента, можно встретить в любом сообществе, группе или же на стене пользователя. Они могут быть представлен в виде фото, видео, анимации или же в текстовом формате.

**Рекламный контент.** Можно встретить как в новостной ленте, так и в пабликах. Реклама может быть представлена в виде изображения с надписью, ссылки и текстовым описанием к ней, анимированного изображения, а также видеоформате. Разместить такой контент может любой пользова-

тель, но зачастую платно или в виде бартера, по предварительной договоренности с администратором ресурса [8].

Новостной контент. Описывает события, которые происходят в мире. Могут быть представлены в любых форматах.

Смешанный контент. В данной группе контент может быть представлен одновременно в различных форматах и описывать одно и то же событие, но различными способами [8].

Изображения. К изображениям относятся фотографии, загружаемые пользователями, картинки, эскизы. Такой вид контента может быть размещен как на странице пользователя, так и в паблике или группе.

Анимации. К анимациям относятся анимированные изображения, длящиеся несколько секунд. Пользователь может отправлять данные изображения своим друзьям, прикреплять их к себе на страницу, сохранять для дальнейшего использования.

Видео. Пользователи могут загружать свои видеоролики в социальную сеть, добавлять в свои видеозаписи уже существующие ролики, делиться ими с друзьями, комментировать, делиться данными видео в сторонних социальных сетях [8,9].

Аудио. Аудио контент является особенностью социальной сети для общения ВКонтакте. Пользователи могут искать различные песни, добавлять их в свои аудиозаписи, создавать альбомы с любимыми треками, просматривать популярные песни, песни своих друзей.

Объявления. Существуют различные паблики, которые размещают посты-объявления. Это может быть покупка или продажа чего-либо, аренда, обмен, поиск [9].

Опросы. Многие сообщества для привлечения внимания пользователей устраивают опросы. Опросы могут быть на любую тематику, скрытые или открытые [8].

Все пользователи в социальной сети могут взаимодействовать между собой.

Создавая пост, пользователь может размещать в нем контент в любом формате: фото, видео, аудио, текст, граффити, опросы. Просматривая новостную ленту, пользователь видит в ней изменения и обновления, которые произошли на страницах его друзей или в пабликах и сообществах, на которые он подписан. Он может или оценить понравившийся ему пост, поставив лайк, или же поделиться им со своими друзьями, сделав репост на свою страницу, в группу, в которой он может являться администратором или в личные сообщения пользователю или группе пользователей [9].

Сообщения в социальной сети Вконтакте позволяют общаться пользователем между собой в приватном режиме. Это означает, что данная переписка будет доступна только сторонам, которые участвуют в беседе. Пользователь, который создал беседу, является её администратором и только он может добавлять и удалять пользователей из данной беседы. Главной особенностью сообщений является возможность отправки звукового сообщения, то есть записи.

В социальной сети ВКонтакте выделяют следующие субъекты: администрация сети, модераторы, редакторы, активные и пассивные пользователи. Администрация сети предоставляет права модераторам и редакторам, а также устанавливает настройки для ресурсов сети и настройки доступа для сторонних социальных сетей. Редакторы отвечают за редактирование статей и контента в группах. Модераторы занимаются управлением групп и проверяют профили пользователей на запрещенный контент, также они могут редактировать статьи. Пассивные пользователи в сети могут лишь ознакомиться с контентом из групп, профилей пользователя и со стены, но только в том случае, если доступ не скрыт настройками приватности. Активные пользователи могут создавать, удалять и редактировать

контент практически во всех ресурсах сети, кроме закладок и документов [8].

На основе выше перечисленных субъектов и объектов сети на рисунке ниже отображена структурно-функциональная схема взаимодействия компонентов сети.

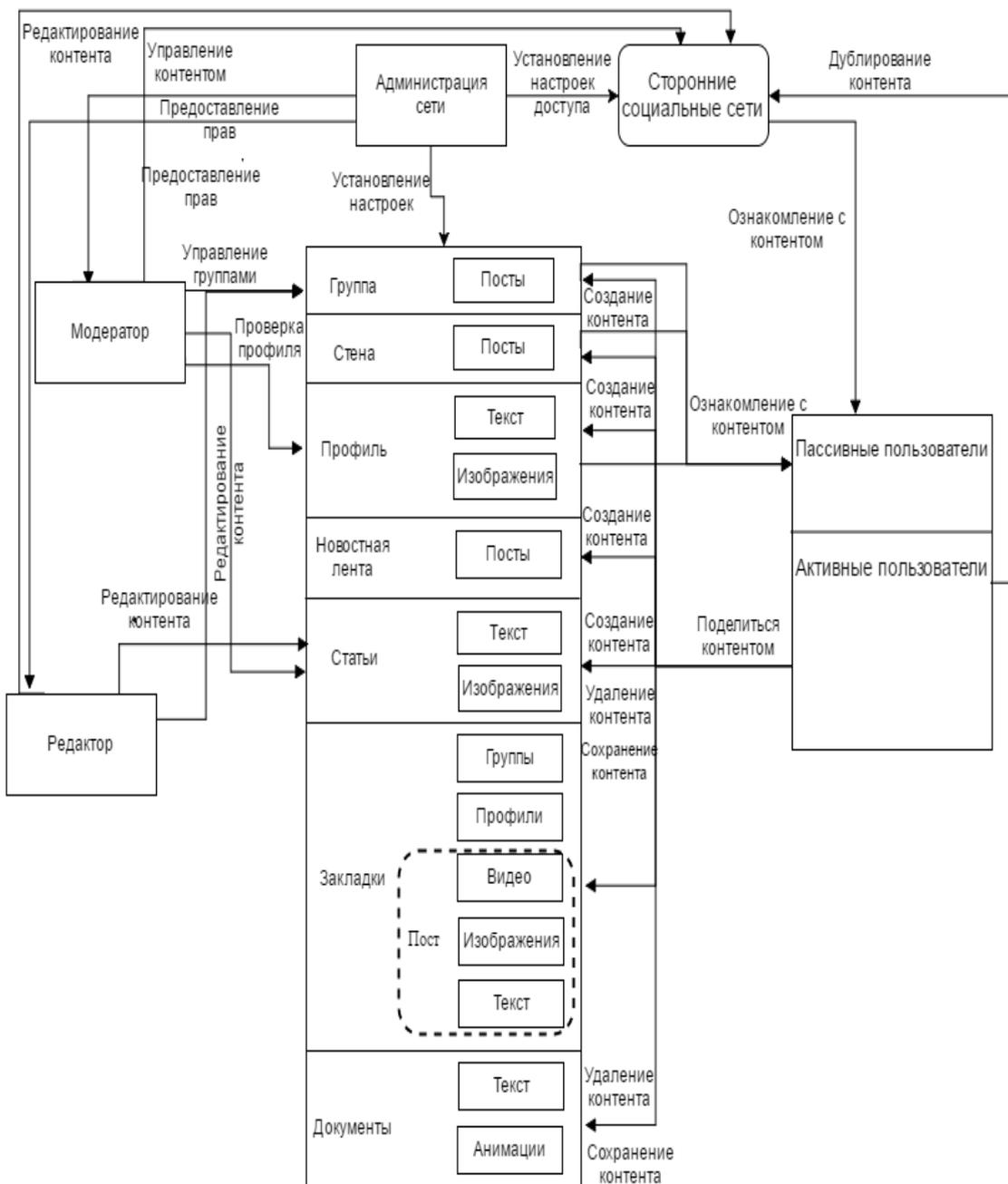


Рис. 4. Структурно-функциональная схема взаимодействия компонентов сети

Данные, полученные в ходе исследования социальной сети для общения ВКонтакте, могут быть базой для анализа и управления информационными рисками пользователей сети при распространении вредоносного контента.

Перейдем к статистическим данным, характеризующим количество узлов (пользователей) и ребер (дружеские связи, основанные на обмене постами, их комментировании и публикации). Данные по социальной сети ВКонтакте представлены в таблице ниже [8,11].

Таблица 1 – Статистические данные социальной сети ВКонтакте

Вершины	1632803
Ребра	30622564
Средняя степень вершины	291,25
Средний коэффициент кластеризации	0.1094
Количество треугольников	32557458
Доля замкнутых треугольников	0.01611
Диаметр	11

Необходимым элементом для анализа социальной сети ВКонтакте является нахождение взвешенных метрик [9,11].

Формат имеющихся статистических данных соответствует трехместному предикату

$$\Gamma(x_i, a_{ij}, x_j) \Leftrightarrow \Gamma(i, \delta(a_{ij}), j),$$

где  $i$  и  $j$  – номера вершин  $x_i$  и  $x_j$  в сети;

$\delta(a_{ij})$  - вес дуги  $a_{ij}$ , связывающей  $x_i$  и  $x_j$ , и направленной от  $i$  к  $j$

Под весом дуги подразумевается динамический ресурс, т.е. передачу определенного объема  $V$  и ценности  $C$  наполнителя сети в единицу времени, т.е.

$$\delta(a_{ij}) = \frac{\partial [cv]}{\partial t} = \langle C \rangle V',$$

где  $s$  усредненной ценностью  $\langle c \rangle$  связывается популярность с различными параметрами пользователей  $i$  и  $j$ , а  $v'$  относится к интенсивности обмена контентом.

Данный формат исходных данных позволяет построить звездную матрицу, элементы строки которой соответствуют дугам, входящим в данную вершину, а элементы столбца – дугам, исходящим из вершины. Очевидно, что матрица будет квадратная, размерности, равной количеству вершин в сети, и ее диагональные элементы будут равны нулю.

Таблица 2 – Звездная матрица социальной сети ВКонтакте

Номер вершины	2	3	4	5	6	...	246	255	292	294	311	315	318	322	326	359	N
2	0	2	3	4	1	...	2	5	2	3	4	5	2	0	0	0	...
3	0	0	1	2	1	...	0	4	0	1	4	4	0	3	0	0	...
4	0	0	0	4	3	...	0	0	0	5	0	3	0	0	0	0	...
5	0	0	0	0	0	...	0	0	0	0	0	1	0	0	0	0	...
6	0	0	0	0	0	...	0	0	0	0	0	0	0	0	0	0	...
7	0	0	0	0	0	...	0	0	0	0	0	0	0	0	0	0	...
8	0	0	0	0	0	...	0	0	0	0	0	0	0	0	0	0	...
9	0	0	0	0	0	...	0	0	0	0	0	0	0	0	0	0	...
10	0	0	0	0	0	...	0	0	0	0	0	0	0	0	0	0	...
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
196	0	0	0	0	0	...	0	0	0	0	0	0	0	0	0	0	...
203	0	0	0	0	0	...	0	0	0	0	0	0	0	0	0	0	...
232	0	0	0	0	0	...	1	1	4	0	0	0	0	5	0	0	...
246	0	0	0	0	0	...	0	0	1	0	0	0	0	0	0	0	...
255	0	0	0	0	0	...	0	0	2	2	3	0	0	0	4	4	...
292	0	0	0	0	0	...	0	0	0	1	3	0	0	0	3	0	...
294	0	0	0	0	0	...	0	0	0	0	0	3	0	0	4	0	...
311	0	0	0	2	0	...	2	0	0	0	0	0	0	0	5	0	...
315	0	0	0	0	0	...	0	0	3	0	0	0	0	0	0	0	...
326	0	0	1	0	0	...	0	0	0	0	0	0	4	0	0	0	...
359	0	0	0	0	0	...	0	0	0	0	0	0	0	0	0	0	...
N	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...

Для последующих процедур моделирования процессов диффузии контента нужно для каждой вершины знать количество исходящих из нее дуг, которые называются степенью ее исхода. Для этого необходимо определить диагональную матрицу степени вершин сети.

Таблица 3 - Диагональная матрица степеней вершин сети ВКонтакте

Номер вершины	2	3	4	5	6	...	246	255	266	292	294	311	315	326	359	N
2	5	0	0	0	0	...	0	0	0	0	0	0	0	0	0	...
3	0	3	0	0	0	...	0	0	0	0	0	0	0	0	0	...
4	0	0	2	0	0	...	0	0	0	0	0	0	0	0	0	...
5	0	0	0	1	0	...	0	0	0	0	0	0	0	0	0	...
6	0	0	0	0	5	...	0	0	0	0	0	0	0	0	0	...
7	0	0	0	0	0	...	0	0	0	0	0	0	0	0	0	...
8	0	0	0	0	0	...	0	0	0	0	0	0	0	0	0	...
9	0	0	0	0	0	...	0	0	0	0	0	0	0	0	0	...
10	0	0	0	0	0	...	0	0	0	0	0	0	0	0	0	...
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
196	0	0	0	0	0	...	0	0	0	0	0	0	0	0	0	...
203	0	0	0	0	0	...	0	0	0	0	0	0	0	0	0	...
232	0	0	0	0	0	...	0	0	0	0	0	0	0	0	0	...
246	0	0	0	0	0	...	5	0	0	0	0	0	0	0	0	...
255	0	0	0	0	0	...	0	3	0	0	0	0	0	0	0	...
266	0	0	0	0	0	...	0	0	4	0	0	0	0	0	0	...
292	0	0	0	0	0	...	0	0	0	5	0	0	0	0	0	...
294	0	0	0	0	0	...	0	0	0	0	1	0	0	0	0	...
311	0	0	0	0	0	...	0	0	0	0	0	5	0	0	0	...
315	0	0	0	0	0	...	0	0	0	0	0	0	4	0	0	...
326	0	0	0	0	0	...	0	0	0	0	0	0	0	5	0	...
359	0	0	0	0	0	...	0	0	0	0	0	0	0	0	3	...
N	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...

Для того, чтобы в дальнейшем осуществить репрезентативную выборку из мультиразмерной сети необходимо знать удельный вес ее вершин и дуг. Для этого построим квадратную матрицы взвешенной центрально-сти элементов сети. Пронормируем веса вершин и дуг по сумме весов всех дуг сети

$$\sum_{i \neq j} \delta(a_{ij}),$$

т.е. – по суммарному трафику сети. Тогда нормированная величина, полученная в данном случае

$$\delta(\bar{a}_{ij}) = \delta(a_{ij}) / \sum_{i \neq j} \delta(a_{ij}),$$

будет показывать удельный вес трафика в дуге  $a_{ij}$  к суммарному трафику сети. Она и будет характеризовать степень ее взвешенной центральности.

Так как суммарный трафик сети не разделяет входящие и исходящие дуги, для определения взвешенной центральности вершины  $x_s$  используем сумму

$$\sum_i \delta(a_{si}) + \sum_j \delta(a_{js}),$$

далее пронормируем ее по суммарному трафику сети. В результате получим нормированную величину

$$\delta(\bar{x}_s) = [\sum_i \delta(a_{si}) + \sum_j \delta(a_{js})] / \sum_{i \neq j} \delta(a_{ij}),$$

которая характеризует удельный вес трафика, проходящего через вершину  $x_s$ , по отношению ко всему трафику сети. Данная величина характеризует степень взвешенной центральности данной вершины. На основе полученных расчетов построим квадратную матрицу взвешенной центральности элементов сети.

Таблица 4 - Квадратная матрица взвешенной центральности элементов сети.

Но- мер вер- шины	2	3	4	5	6	...	255	266	292	294	311	315	326	N
2	0,09	0,002	0,002	0,0031	0,003	...	0,005	0,0004	0,0004	0,004	0,001	0,004	0	...
3	0	0	0,003	2	0,003	...	0,004	0,0032	0	0,003	0,001	0,002	0	...
4	0	0	0	4	3	...	0	0	0	0,004	0	0,002	0	...
5	0	0	0	0	0	...	0	0	0	0	0	0,003	0	...
6	0	0	0	0	0	...	0	0	0	0	0	0	0	...
7	0	0	0	0	0	...	0	0	0	0	0	0	0	...
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
311	0	0	0	0,0036	0	...	0	0	0	0	0	0	0,005	...
315	0	0	0	0	0	...	0	0	0	0	0	0	0	...
326	0	0	0,004	0	0	...	0	0	0	0	0	0	0	...
N	...	...	...	...	...	...	...	...	...	...	...	...	...	...

Мультиразрядность социальных сетей не позволяет использовать весь объем данных для исследования социальных сетей. Поэтому необходима репрезентативная выборка, которая имеет значительно меньшую размерность, но с заданной точностью сохраняет свойства исходной сети.

Тривиальный подход заключается в отборе вершин по их степени, т.е. количеству инцидентным им дуг. Однако, такой критерий явно не учитывает тот факт, что каждая дуга имеет свой вес, т.е. отсутствует взвешенная оценка с точки зрения трафика.

Поэтому уместно использование квадратной матрицы взвешенной центральности элементов сети, а точнее – ее диагонали, где определена степень взвешенной центральности каждой вершины.

Такой подход наиболее объективен, ибо учитывает трафик – основу функционирования любой сети.

Для получения репрезентативной выборки воспользуемся алгоритмом, реализованным совместно с партнерами по комплексной работе.

Рассмотрим диагональные элементы матрицы взвешенной центральности, полученной выше. Далее произведем сортировку вершин, по убыванию оценки взвешенной центральности. Будем считать, что допустима потеря не более 5% трафика для усеченной модели сети. На основе данного полученного усечения построим следующую матрицу.

Таблица 5 - Проранжированная усеченная матрица социальной сети ВКонтакте

Номер вершины	6521	19562	29481	...	44323	5638	...	27154	27219	830	...	19063	N
6521	0	0	0	...	0	0	...	0	0	0	...	0	0
19562	0	0	0,0003 3691	...	0	0	...	0	0	0	...	0	0
29481	0	0,000 3265	0	...	0,0002 6941	0	...	0,0001 5528	0	0	...	0	0
...	...	...	...	...	...	...	...	...	...	...	...	...	...
44323	0	0	0	...	0	0	...	0	0	0	...	0,0001 3957	0
13596	0,000 2601	0	0	...	0	0,00032 0649	...	0	0	0	...	0	0
5638	0	0	0	...	0	0	...	0	0	0,0002 6597	...	0	0
...	...	...	...	...	...	...	...	...	...	...	...	...	...
27154	0	0	0	...	0	0	...	0	0	0	...	0	0
27219	0	0	0	...	0	0	...	0	0,00015 2832	0	...	0	0
830	0	0	0	...	0	0	...	0	0	0	...	0	0
...	...	...	...	...	...	...	...	...	...	...	...	...	...
19063	0	0	0	...	0	0	...	0	0	0	...	0	0
N	0	0	0	...	0	0	...	0	0	0	...	0	0

На основе данной матрицы, с помощью программного обеспечения Gephi [10] построим усеченный граф социальной сети:

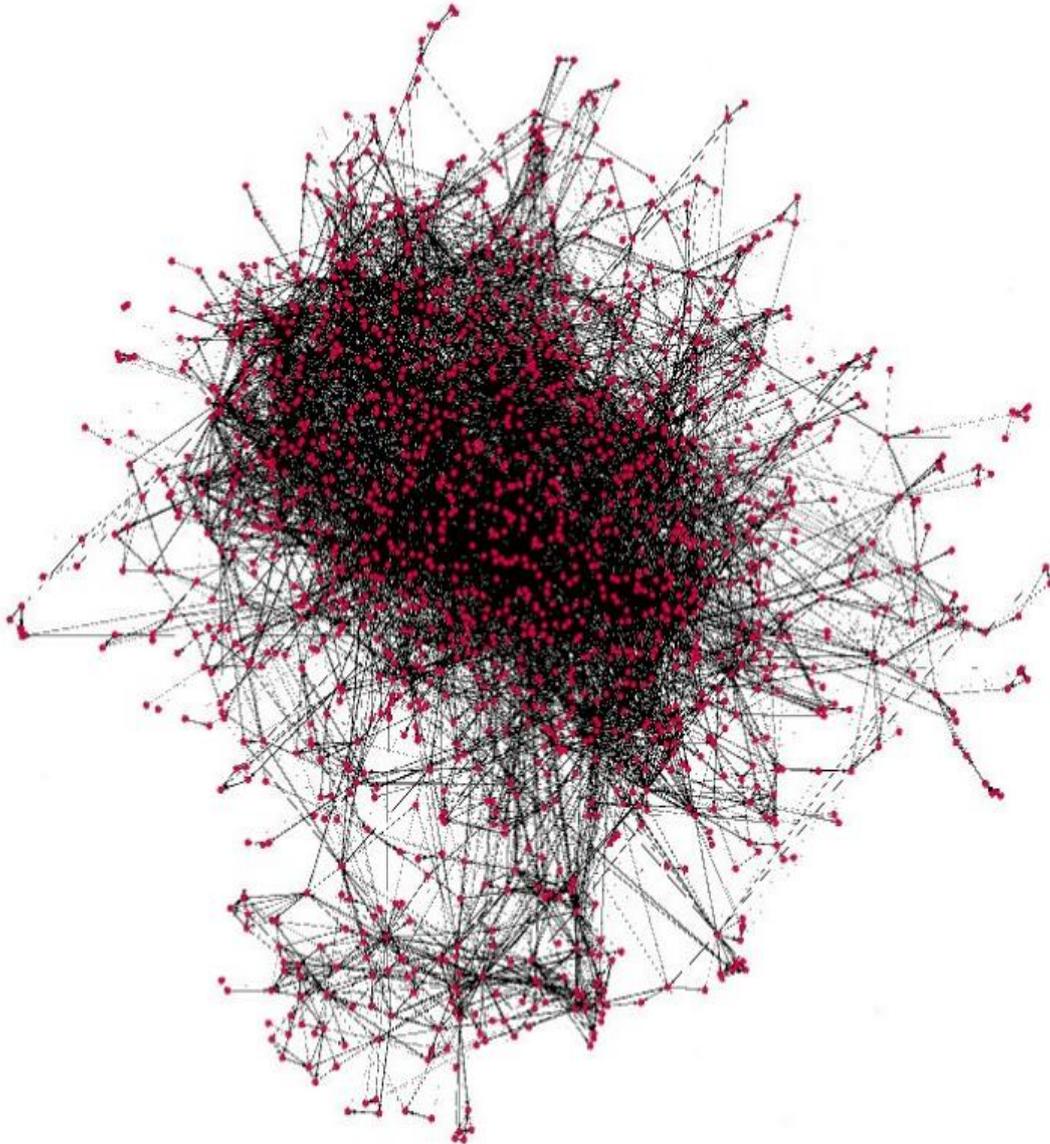


Рис. 5. Визуализация репрезентативной выборки

Последний этап алгоритма состоит в формировании итоговой матрицы в виде, удобном для последующего моделирования информационной диффузии. Отсюда определяется ожидаемое количество вирусованных контентом элементов.

Таблица 6 - Модифицированная матрица смежности социальной сети

ВКонтакте

Номер вершины	6521	19562	29481	...	44323	13596	5638	...	27154	27219	830	...	19063	N
6521	0	0	0	...	0	0	0	...	0	0	0	...	0	0
19562	0	0	1	...	0	0	0	...	0	0	0	...	0	0
29481	0	1	0	...	1	0	0	...	1	0	0	...	0	0
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
44323	0	0	0	...	0	0	0	...	0	0	0	...	-1	0
13596	-1	0	0	...	0	0	-1	...	0	0	0	...	0	0
5638	0	0	0	...	0	0	0	...	0	0	1	...	0	0
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
27154	0	0	0	...	0	0	0	...	0	0		...	0	0
27219	0	0	0	...	0	0	0	...	0	-1	0	...	0	0
830	0	0	0	...	0	0	0	...	0	0	0	...	0	0
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
19063	0	0	0	...	0	0	0	...	0	0	0	...	0	0
N	0	0	0	...	0	0	0	...	0	0	0	...	0	0

Для доказательства репрезентативности выборки необходимо рассчитать среднее квадратическое отклонение исходных данных и полученной выборки.

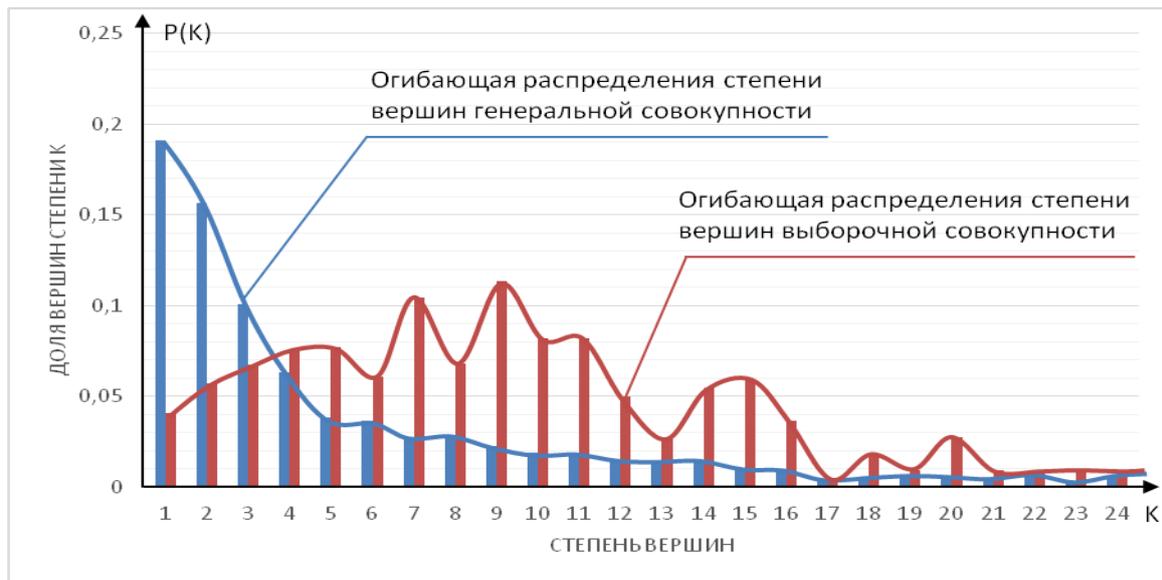


Рис. 6. Распределение количества вершин по степеням

Для того, чтобы доказать репрезентативность выборки, воспользуемся критерием Пирсона. На основе доказательства о принадлежности полученной выборки сети к степенному закону распределения, рассчитаем критерий Пирсона [12].

$$\chi_{\text{набл}}^2 = \sum_{i=1}^k \frac{(n - np_i)^2}{np_i},$$

где  $n$  – объем выборки;

$p_i$  - вероятность попадания случайной величины  $X$ .

Следуя представленной формуле, рассчитаем значение  $\chi^2$ . В данном случае  $\chi^2=15.528$ . Следовательно,  $n = 14$  (выборка) – 1 (количество параметров в плотности степенного распределения) = 13 (число степеней свободы).

Далее выберем уровень значимости  $\alpha = 0.05$ , тогда  $p = 1 - \alpha = 0.95$ . Табличное значение  $\chi^2$  равно 22.4. Получается, что табличное значение  $\chi^2$  больше вычисленного. Следовательно, гипотеза верна, и приведенное степенное распределение согласуется с результатами эксперимента, что и требовалось доказать.

Для выполнения главного и необходимого свойства выборки, рассчитаем среднеквадратичное отклонение (СКО) усеченной и генеральной выборок и коэффициент корреляции [9].

$$\sigma_{XY} = \sqrt{\sum_{i=1}^n (X_i - Y_i)^2},$$

$$\text{cov}(X, Y) = \frac{1}{n} \sum_{t=1}^n (X_t - \bar{X})(Y_t - \bar{Y}),$$

$$\sigma_X = \sqrt{\frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^2},$$

$$r_{XY} = \frac{\text{cov}(X, Y)}{\sigma_X \sigma_Y},$$

где:  $n$  – общее число вершин;

$X_i$  –  $i$ -ый элемент выборки;

$\bar{X}$  – среднее арифметическое выборки;

$\sigma_X, \sigma_Y$  – среднеквадратическое отклонение для совокупностей  $X$  и  $Y$  соответственно;

$\sigma_{XY}$  – среднеквадратическое отклонение значений выборочной совокупности от генеральной;

$cov(X, Y)$  – коэффициент ковариации совокупностей  $X$  и  $Y$ ;

$r_{XY}$  – коэффициент корреляции.

Запишем рассчитанные результаты (табл. 7).

Таблица 7 – Полученные значения сравнения совокупностей

Коэффициент	Значение
СКО	0,06746
Коэффициент корреляции	0,75308

Исходя из полученных данных, можно сделать вывод, что имеет место репрезентативная выборка, так как значения являются допустимыми.

В ходе работы были проведены исследования социальной сети для общения ВКонтакте. Опираясь на передовые исследования отечественных и зарубежных ученых [1-7], были выполнены алгоритмы преобразования исходных данных сети и нахождения репрезентативной выборки. На основе полученных данных была построена структурно-функциональная модель сети со всеми её особенностями и были получены данные, описывающие распространение трафика в данной социальной сети.

Литература:

1. Analytical estimation of the component viability of distribution automated information data system / G.A. Ostapenko, D.G. Plotnicov, O.Y. Makarov, N.M. Tikhomirov, V.G. Yurasov // World Applied Sciences Journal. – 2013. – 25 (3). – P. 416-420.
2. Analytical models of information-psychological impact of social information networks on users / G.A. Ostapenko, L.V. Parinova, V.I. Belonozhkin, I.L. Bataronov, K.V. Simonov // World Applied Sciences Journal. – 2013. – 25 (3). – P. 410-415.
3. Assessment of the system's EPI-resistance under conditions of information epidemic expansion / N.M. Radko, A.G. Ostapenko, S.V. Mashin, O.A. Ostapenko, D.V. Gusev // Biosciences Biotechnology Research Asia. – 2014. – Vol. 11 (3). – P. 1781-1784.
4. Discreet risk-models of the process of the development of virus epidemics in non-uniform networks / V.V. Islamgulova, A.G. Ostapenko, N.M. Radko, R.K. Babadzhanov, O.A. Ostapenko // Journal of Theoretical and Applied Information Technology. – 2016. – P. 306-315.
5. Flood-attacks within the hypertext information transfer protocol: damage assessment and management / A.G. Ostapenko, M.V. Bursa, G.A. Ostapenko, D.O. Butrik // Biosciences Biotechnology Research Asia. – 2014. – Vol. 11 (Spl.End). – P. 173-176.
6. Optimization of expert methods used to analyze information security risk in modern wireless networks / S.A. Ermakov, A.S. Zavorykin, N.S. Kolenbet, A.G. Ostapenko, A.O. Kalashnikov // Life Science Journal. – 2014. – № 11(10s). – P. 511-514.
7. Peak risk assessing the process of information epidemics expansion / N.M. Radko, A.G. Ostapenko, S.V. Mashin, O.A. Ostapenko, A.S. Avdeev // Biosciences Biotechnology Research Asia. – 2014. – Vol. 11 – P. 251-255.

8. Блог социальной сети ВКонтакте. – Электрон. Дан. – Режим доступа: <https://vk.com/blog>.
9. Группа LIVE. Всё о социальной сети ВКонтакте. – Электрон. Дан. – Режим доступа: <https://vk.com/live>.
10. Средство визуализации данных. - Электрон. Дан. – Режим доступа: <https://gephi.org>.
11. Статистические данные социальной сети ВКонтакте. – Электрон. Дан. – Режим доступа: <https://vk.com>.
12. Теория вероятностей и математическая статистика / Гмурман В.Е. // Учебное пособие. Высшее образование. – 2007. – 290 с.

Воронежский научно-образовательный центр  
управления информационными рисками

**УДК 004.056.57**

**СТРУКТУРНО – ФУНКЦИОНАЛЬНАЯ СПЕЦИФИКА  
РАСПРОСТРАНЕНИЯ КОНТЕНТА В СОЦИАЛЬНОЙ  
FOURSQUARE**

**С.И. Уколов, А.В. Паринов**

Геолокационный сервис расширяет границы «социальности». Раньше индивид мог демонстрировать окружающим свою личность через страницу в социальной сети, где значились его: интересы, любимая музыка, места, которые ему нравятся, фотографии. Это была «статичная» информация, как слепок с портрета пользователя. Застывшие факты, запечатленные моменты прошлого [9, 17].

Эту картинку разработчики «оживили» сервисом микроблогов и живой лентой мест на карте. Внедрение сервиса геолокаций стало логичным продолжением этой «актуализации» - пользователь обрел «динамичность». Теперь он может рассказать больше о своем образе жизни, распорядке дня, социальном уровне. Как любой потребитель бренда, он перенимает часть декларируемого брендами успеха. В данном случае - брендов заведений, сферы услуг и мест отдыха [12].

В этом смысле успех геолокационных сервисов можно объяснить тем, что они фактически восполнили незанятую нишу «потребления брендов»: если раньше все демонстрировали свой статус брендами-продуктами, то теперь появилась возможность добавить к этому подчас более показательные бренды сферы обслуживания [11].

Хорошо продуманная социальная механика, обеспечивающая высокую вовлеченность пользователей в процесс рекомендаций, оценок, ранжирования.

Практически все успешные социальные геолокационные сервисы в той или иной мере используют игровую форму взаимодействия с пользователем, чтобы сохранить его мотивацию для самостоятельной генерации контента. В этом большое достижение сервисов геолокации - они не завязаны на модерации контента администраторами и не обременены необходимостью наполнения объемного массива данных по локациям [9].

Социальные сети для отзывов и обзоров предоставляют информацию о конкретном месте для посещения или путешествия, а также о продуктах, используемых в повседневной жизни. Любой пользователь сети сможет найти необходимую ему информацию или добавить свой отзыв. Актуальностью этих сетей является тот факт, что люди все больше основываются на мнениях других людей при выборе чего-либо, чем рекламе в СМИ [8].

В Foursquare место является физическим местоположением. Это может быть магазин, ресторан, железнодорожная станция, кинотеатр, или частная резиденция.

Создание места: Пользователи могут создавать центры через веб-сайта или мобильное приложение Foursquare. Каждому месту присваивается уникальный идентификатор на карте. При создании места, пользователю предлагается представить несколько атрибутов, такие как имя, адрес встречи, местоположение, категория, почтовый индекс, улица и т.д. Адрес состоит из номера и названия улицы, а место расположения задается пользователем, установив отметку на карте в Foursquare. Пользователи выбирают категории для места из списка, определяемого Foursquare [8, 10].

Foursquare определяет трехуровневую иерархическую структуру: категории, теги, свойства. Есть девять категорий верхнего уровня: искусство и развлечения, колледж и университет, еда, профессиональные и другие места, ночная жизнь, природа и отдых, магазины и сервисы, путешествия и транспорт, резиденции. Посмотрев на популярность предложенных сетью

категорий, можно сказать, что в основном пользователи данной сети интересуются местами общественного питания (рис. 1).

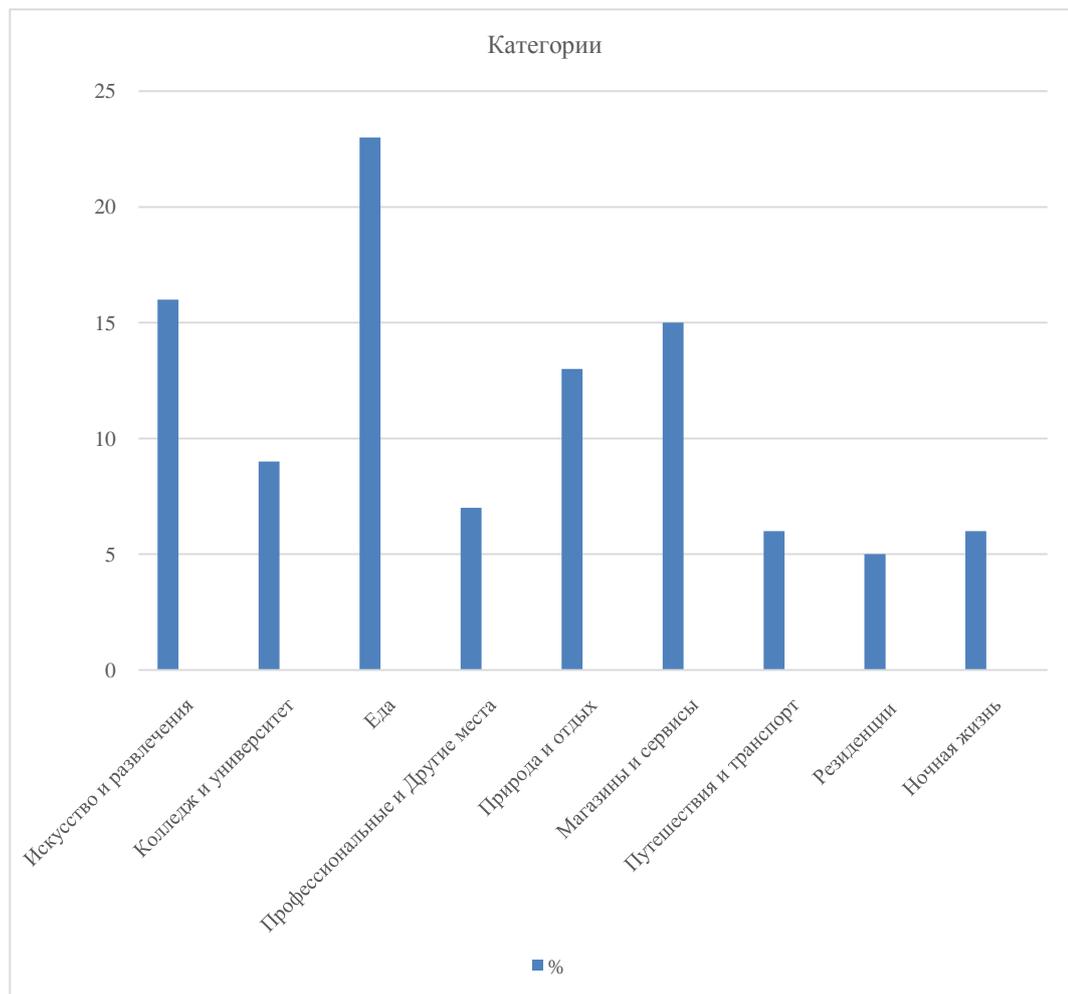


Рис. 1. Категории верхнего уровня

Пользователи должны указать категорию при создании места. Foursquare запустило особенность категории 10-го марта 2010 года, так что места созданные до этой даты, не имеют какую-либо информацию о категории. Особый случай представляет "Home (Private)" категория, которая является подкатегорией жительство. Места в этой категории являются конфиденциальной информацией; Кроме того, Foursquare гарантирует, что детали домашнего (частного) места будут видны только владельцу [8, 9, 11]. А другие пользователи будут видеть это место в уменьшенном виде на

карте вместо точного местоположение места встречи. Пользователи Foursquare могут явно выражать свои интересы к местам двумя способами: комментарии и система оценок [10]. Рассмотрим структурно-функциональную схему сети Foursquare (рис. 2).

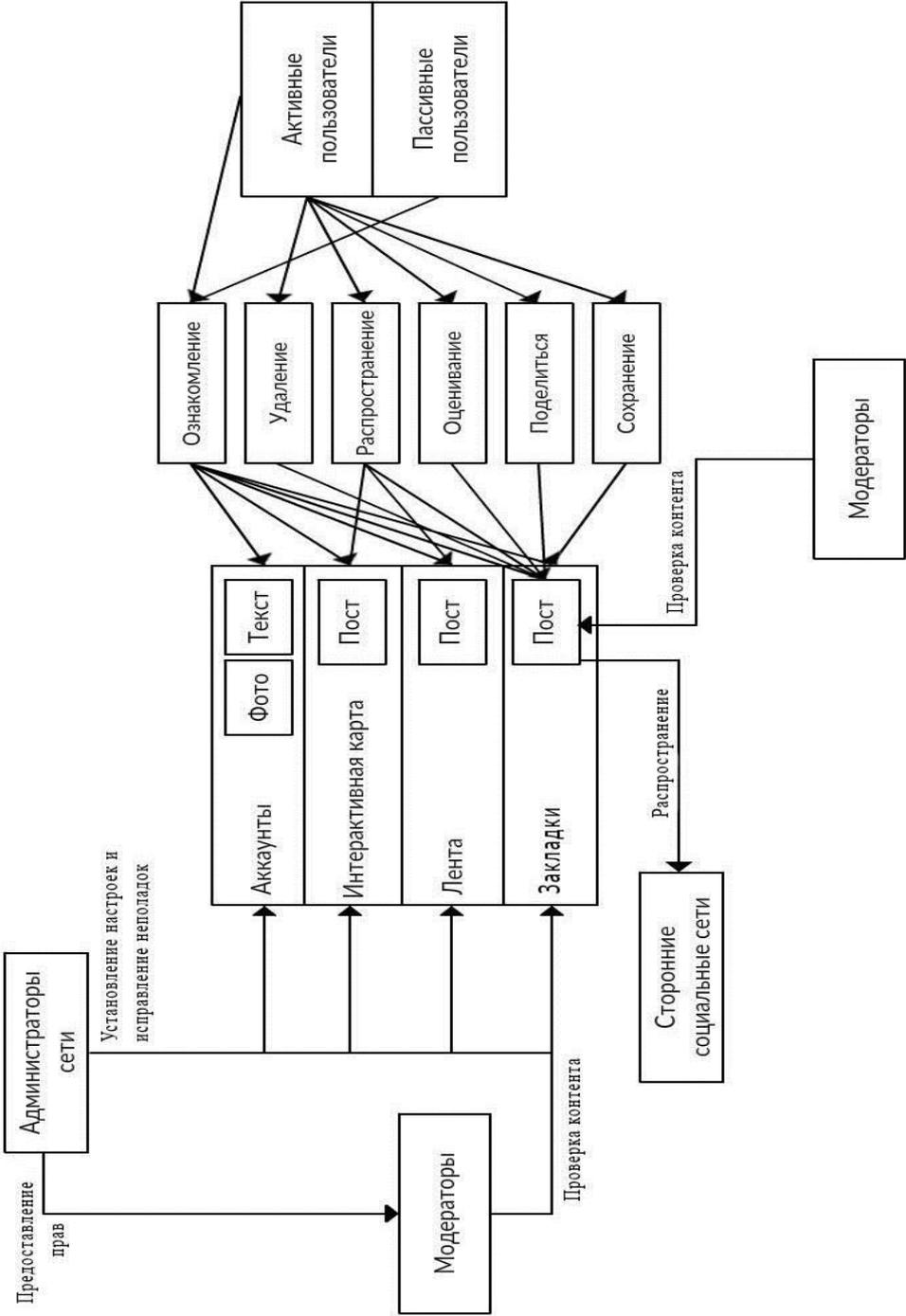


Рис. 2. Структурно-функциональная схема

В первую очередь следует обратить внимание на две основные характеристики Foursquare как интернет-ресурса: во-первых, оно мобильно, то есть, оно предполагает свое использование в мобильном телефоне, несмотря на то, что также может быть запущено и на обычном компьютере). Во-вторых, оно организовано в виде социальной сети, значит, оно позволяет пользователям регистрировать собственные отдельные профили и вступать в некие виртуальные взаимоотношения (добавлять в закладки) – становится так называемыми «друзьями» [12]. Основой Foursquare является все та же географическая карта, представленная в цифровом виде под названием OpenStreetMap, отражающая всю ту же «внешнюю» по отношению к интернету реальность, но фактически не принадлежащая никакому конкретному юридическому субъекту, в отличие от Google Maps. Тем самым обнаруживается, что изначальная основа этой карты (сама репрезентация субъектов географического пространства) существует в Foursquare «по умолчанию» по принципам свободного распространения информации, то есть то самое «рациональное» знание уже обнаруживает свое существование посредством существования такого свободного программного обеспечения. Однако, наибольший интерес представляет то, каким образом сами пользователи в рамках сети могут пользоваться возможностью свободно изменять это знание внутри сети [8, 15].

Тут в первую очередь не следует забывать о том, что сам Foursquare является субъектом современного капитализма, существующем в рамках социальной сети (чьей целью является привлечение наибольшего количества пользователей, их аккумуляция) и именно посредством этого субъекта пользователи и осуществляют возможное «создание» самого знания (которым является определенный артефакт картографии). Другими словами, пользователи словно платят своим присутствием внутри данной социальной сети за возможность участвовать в создании определенного знания по-

средством этой же сети [9]. Именно в этом состоит парадокс создания знания в целом (и картографического знания в частности) в эпоху “Web 2.0”. Знание само по себе кажется «свободным» - оно не принадлежит никому, но за доступ к нему пользователю приходится интегрироваться в деятельность того или иного крупного субъекта капитализма, причем платить этому субъекту он будет одним своим присутствием. Полученное знание сложно будет назвать именно «рациональным», так как пользователи сами по себе юридически не наделены ответственностью создавать исключительно «истинные» по отношению к «внешнему» пространству картографические объекты внутри самого Foursquare [14, 17].

Размышляя же о пользовательских возможностях Foursquare (рис. 3), как о медиа цифровой эпохе, позволяющих пользователю участвовать в процессе картографии, в первую очередь не стоит забывать о том, что сам по себе Foursquare является социальной сетью, предполагающей участие большого количества пользователей в создании самого контента.

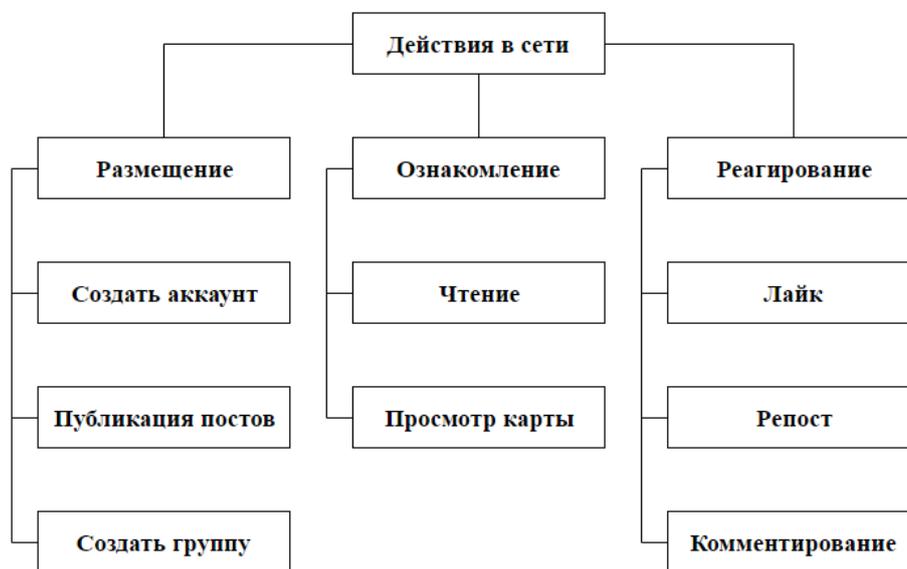


Рис. 3. Классификация действий в сети

«Объекты» Foursquare сами по себе динамичны; именно они создают карту внутри этого ресурса, репрезентируя своим существованием внутри

Foursquare «внешнюю» по отношению к нему реальность. Динамичными же они являются потому, что возможностью создания этих объектов обладают и сами пользователи [13].

К примеру, на Foursquare существуют и такие «объекты», которые являются «значимыми» абсолютно для всех пользователей Foursquare (к примеру – достопримечательности, известные рестораны, места отдыха и т.д.), которые были перенесены Foursquare с проекта OpenStreetMap и объекты, существующие исключительно внутри Foursquare и созданные самими отдельными пользователями (эти «объекты», таким образом, становятся значимыми лишь для определенной подгруппы пользователей) [8, 16].

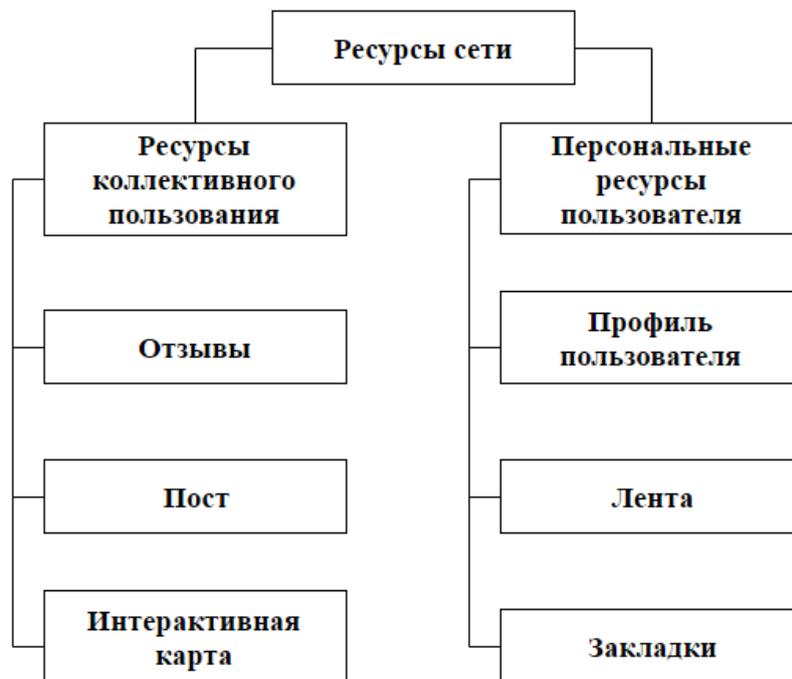


Рис. 4. Классификация объектов сети

Особенностью классификаций объектов среди сетей отзывов и обзоров (рис. 4) является наличие закладок у пользователей, которыми они могут обмениваться, подписавшись друг на друга [14, 15].



Рис. 5. Классификация субъектов сети

Субъекты сети делятся на «Активных» и «Пассивных» (рис. 5). Нам же интересуют «Активные» так как они являются наиболее влиятельными распространителями контента [16].

Администраторы сети задают некие стандарты для контента.

Модератор проверяет правильность введенных данных пользователем.

Генерирующие контент пользователи добавляют места на карте с их описанием.

Распространяющие контент пользователи используют уже имеющийся контент.

Системный администратор контролирует правильность работы БД.

Перейдем к статистическим данным, характеризующим количество узлов (пользователей) и ребер (дружеские связи, основанные на обмене за-

кладками, комментировании статей и их публикации). Данные, изложенные на сайте github по сети Foursquare представлены ниже (табл. 1) [13].

Таблица 1 – Статистические данные социальной сети Foursquare

Характеристические данные	Показатели
Количество узлов	91500
Количество ребер	1681150
Диаметр	25
Средняя длина пути	3,75
Коэффициент кластеризации (%)	1.154
Средний диаметр сети	4,8967

На рисунке ниже представлен граф полученный на основе исходных данных, на нём явно прослеживается кластерность (рис. 6). Следствием является высокий коэффициент кластеризации. Это связано с тем, что пользователи сети в основном подписываются на места в своём городе.

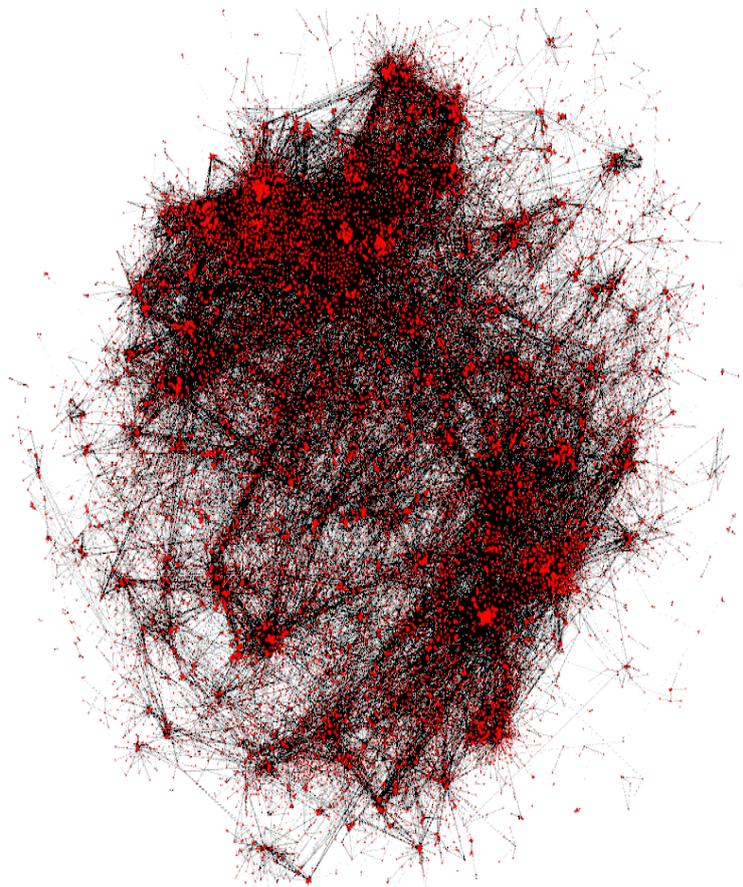


Рис. 6. Визуализация исходного графа

Необходимым элементом для анализа социальной сети Foursquare является нахождение взвешенных метрик.

Статистические данные социальной сети Foursquare, собранные в репозитории Github foursquare-user-dataset, можно представить в виде трехместного предиката, в котором на первом месте располагается узел сети, из которого исходит ребро, подтверждающее связь между пользователями сети, на втором месте – узел, в который входит это же ребро, и на третьем – вес ребра, отражающий количество отзывов и обзоров в единицу времени [13]. По своей структуре они соответствуют трехместному предикату (1):

$$\Gamma(x_i, a_{ij}, x_j) \Leftrightarrow \Gamma(i, \delta(a_{ij}), j), \quad (1)$$

где  $i$  и  $j$  – номера вершин  $x_i$  и  $x_j$  в сети;

$\delta(a_{ij})$  - вес дуги  $a_{ij}$ , связывающей  $x_i$  и  $x_j$ , и направленной от  $i$  к  $j$

Авторы этого формата подразумевают под весом дуги динамический ресурс, т.е. количество отзывов и обзоров определенного объема  $V$  и ценности  $C$  наполнителя сети в единицу времени (месяц) (2):

$$\delta(a_{ij}) = \frac{\partial [cv]}{\partial t} = \langle C \rangle V', \quad (2)$$

где с усредненной ценностью  $\langle C \rangle$  они связывают популярность и др. параметры пользователей  $i$  и  $j$ , а  $V'$  относят к интенсивности обмена контентом.

Исследование сети с подобным количеством узлов и взаимосвязей не предоставляется возможным. Для дальнейшего анализа сети воспользуемся алгоритмом преобразования исходных данных.

Для этого построим звездную матрицу (табл. 2), элементы строки которой соответствуют дугам, входящим в данную вершину, а элементы столбца – дугам, исходящим из вершины. Матрица очевидно будет квадратная, размерности, равной количеству вершин в сети, и ее диагональные элементы будут равны нулю.

Таблица 2 - Звездная матрицу взвешенной связности исходной сети

Номер вершины	1	2	3	4	5	...	566	567	568	...	91500
1	-	20	2	0	1	...	0	2	1	...	0
2	1	-	0	4	1	...	0	18	1	...	3
3	9	0	-	2	1	...	0	0	0	...	0
4	0	1	4	-	3	...	0	0	0	...	0
5	4	0	0	0	-	...	0	4	0	...	2
...	...	...	...	...	...	...	...	...	...	...	...
566	0	0	0	0	0	...	-	4	0	...	0
567	0	0	3	0	0	...	0	-	0	...	0
568	0	0	0	0	0	...	0	0	-	...	0
...	...	...	...	...	...	...	...	...	...	...	...
91500	2	0	0	0	1	...	0	0	0	...	-

На диагонали матрицы оставляем прочерк, это означает, что в рассматриваемом графе нет петель. На последующих этапах запишем в эти ячейки значения взвешенных центральностей вершин.

Для моделирования процессов диффузии контента необходимо рассмотреть диагональную матрицу степеней вершин. Заметим, что степенью вершины будем называть количество исходящих из вершины дуг, без учета их веса (табл. 3).

Таблица 3 – Диагональная матрица степеней вершин

Номер вершины	1	2	3	4	5	6	...	566	567	568	...	91500
1	15	0	0	0	0	0	...	0	0	0	...	0
2	0	7	0	0	0	0	...	0	0	0	...	0
3	0	0	2	0	0	0	...	0	0	0	...	0
4	0	0	0	18	0	0	...	0	0	0	...	0
5	0	0	0	0	10	0	...	0	0	0	...	0
6	0	0	0	0	0	9	...	0	0	0	...	0
...	...	...	...	...	...	...	...	...	...	...	...	...
566	0	0	0	0	0	0	...	2	0	0	...	0
567	0	0	3	0	0	0	...	0	13	0	...	0
568	0	0	0	0	0	0	...	0	0	20	...	0
...	...	...	...	...	...	...	...	...	...	...	...	...
91500	0	0	0	0	0	0	...	0	0	0	...	1

Для последующей репрезентативной выборки из мультиразмерной сети необходимо знать удельный вес ее вершин и дуг. Для этого предлагается нормировка их весов по сумме весов всех дуг сети (3)

$$\sum_{\substack{i,j \\ i \neq j}} \delta(a_{ij}), \quad (3)$$

т.е. – по суммарному трафику сети. Тогда нормированная величина, полученная в данном случае будет иметь вид (4)

$$\delta(\bar{a}_{ij}) = \delta(a_{ij}) / \sum_{\substack{i,j \\ i \neq j}} \delta(a_{ij}), \quad (4)$$

она будет показывать удельный вес трафика в дуге  $a_{ij}$  к суммарному трафику сети и будет характеризовать степень ее взвешенной (по трафику) центральности.

Следует заметить, что суммарный трафик сети не разделяет входящие и исходящие дуги. Поэтому для определения взвешенной центральности вершины  $x_s$  можно использовать сумму (5)

$$\sum_i \delta(a_{si}) + \sum_j \delta(a_{js}), \quad (5)$$

которую далее следует пронормировать по суммарному трафику сети. В результате получим нормированную величину (6)

$$\delta(\bar{x}_s) = [\sum_i \delta(a_{si}) + \sum_j \delta(a_{js})] / \sum_{\substack{i,j \\ i \neq j}} \delta(a_{ij}), \quad (6)$$

которая будет характеризовать удельный вес трафика, проходящего через вершину  $x_s$ , по отношению ко всему трафику сети. Это и будет характеризовать степень взвешенной (по трафику) центральности данной вершины.

Следуя вышеизложенному, можно получить квадратную матрицу взвешенной центральности элементов исходной сети (табл. 4).

Таблица 4 – Квадратная матрица взвешенной центральности элементов сети

Номер вершины	1	2	...	87	...	201	202	91500
1	0,0005	0,00052	...	0	...	0	0	0
2	0	0,00423	...	0	...	0	0	0
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
87	0	0	...	0,001029	...	0	0	0
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
201	0	0	...	0	...	0,01235	0	0
202	0	0	...	0	...	0	0,001235	0
91500	0	0	...	0		0	0	0,00018

Для получения репрезентативной выборки воспользуемся алгоритмом, реализованным совместно с партнерами по комплексной работе. За допустимую погрешность примем не более 5% трафика для усеченной модели сети.

Алгоритм получения репрезентативной выборки:

5. Вычленим из исходной матрицы взвешенной центральности элементов сети ее диагонали. Оставшаяся усеченная матрица с обнуленными диагональными элементами будет использована на шаге 4.

6. Отсортируем полученную матрицу взвешенной центральности по убыванию значений степени взвешенной центральности вершин  $\delta(\bar{x}_s)$ , расположенных на главной диагонали.

7. Найдем минимальную сумму значений  $\delta(\bar{x}_s)$ , такую что

$$\sum \delta(\bar{x}_s) > 0.95 \text{ и исключим остальные значения.}$$

8. Удалим из усеченной матрицы все элементы, относящиеся к исключенным на предыдущем шаге путём обнуления столбцов и строк.

Применив данный алгоритм к построенной матрице взвешенной центральности, построим проранжированную усеченную матрицу социальной сети Foursquare (табл. 5):

Таблица 5 – Проранжированная усеченная матрица сети Foursquare

Номер вершины	107	36	208	...	786	...	8780	20	608
107	0	0,00212	0	...	0	...	0	0	0
36	0	0,00423	0	...	0	...	0,000256	0	0
208	0,00123	0	0,01361	...	0	...	0	0	0
⋮	⋮	⋮	⋮	↘	⋮	⋮	⋮	⋮	⋮
786	0	0	0	...	0	...	0	0	0,0000585
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
8780	0	0	0	...	0	...	0	0	0
20	0	0	0	...	0	...	0	0	0
608	0	0	0	...	0	...	0	0	0

Последний этап алгоритма состоит в формировании итоговой матрицы в виде, удобном для последующего моделирования информационной диффузии. Данная матрица выглядит следующим образом (табл. 6):

- недиагональные элементы равны  $\pm 1$ , так как интересует лишь факт смежности вершин истоков (+1) и стоков (-1). В некоторой степени это возврат к формату матрицы смежности;

- диагональные элементы имеют вес  $k$ , равный сумме элементов (+1) столбца.

Таблица 6 - Модифицированная матрицей смежности социальной сети Foursquare

Номер вершины	107	36	208	...	5786	453	...	8780	202	608
107	63	0	1	...	0	1	...	0	0	0
36	0	46	0	...	0	0	...	1	0	0
208	0	0	26	...	0	0	...	0	0	0
⋮	⋮	⋮	⋮	↘	⋮	⋮	⋮	⋮	⋮	⋮
5786	-1	0	0	...	22	0	...	0	0	0
453	0	0	0	...	0	18	...	0	1	0
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
8780	0	0	-1	...	0	0	...	6	0	0
202	0	0	0	...	0	0	...	0	12	0
608	0	0	0	...	0	0	...	0	0	5

Получим оценку параметров графа выборки с помощью программного обеспечения Gephi (табл. 7) [13].

Таблица 7 - Параметры графа выборки

Параметры графа	Выборка
Средняя степень	9,153
Средняя взвешенная степень	14,019
Диаметр графа	12
Средняя длина пути	3,88
Количество ребер	2398
Количество узлов	608

На основе проранжированной усеченной матрицы сети, с помощью программного обеспечения Gephi построим усеченный граф (рис. 7) [12]:

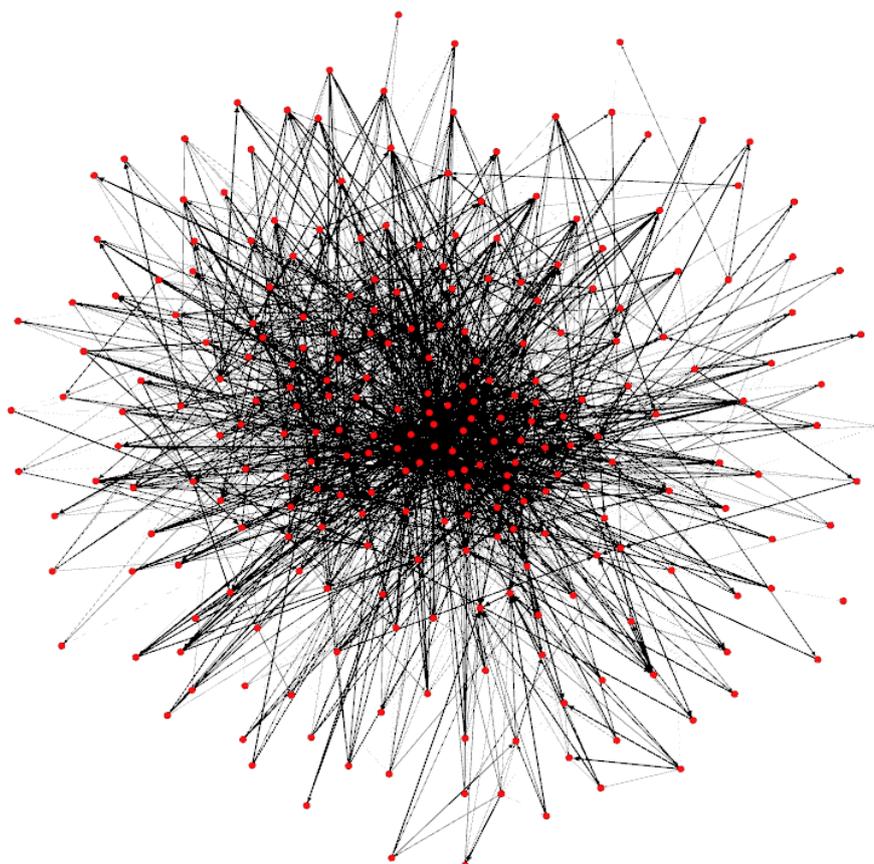


Рисунок 7 - Визуализация полученной выборки

На (рис. 7) видно, что на получившейся вследствие применения алгоритма выборки, граф сохранил схожую структуру с исходным графом.

Для окончательного ответа на вопрос о репрезентативности выборки вычислим сравнительные параметры, описывающие связность значений исходного распределения долей и значений выборочной совокупности. Рассчитаем значения среднеквадратичного отклонения (СКО) по формуле (7) значений двух совокупностей и коэффициент корреляции (8-10).

$$\sigma_{XY} = \sqrt{\sum_{i=1}^n (X_i - Y_i)^2}, \quad (7)$$

$$cov(X, Y) = \frac{1}{n} \sum_{t=1}^n (X_t - \bar{X})(Y_t - \bar{Y}), \quad (8)$$

$$\sigma_X = \sqrt{\frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^2}, \quad (9)$$

$$r_{XY} = \frac{cov(X, Y)}{\sigma_X \sigma_Y}, \quad (10)$$

где:  $n$  – общее число вершин;

$X_i$  –  $i$ -ый элемент выборки;

$\bar{X}$  – среднее арифметическое выборки;

$\sigma_X, \sigma_Y$  – среднеквадратичное отклонение для совокупностей

$X$  и  $Y$  соответственно;

$\sigma_{XY}$  – среднеквадратичное отклонение значений выборочной совокупности от генеральной

$cov(X, Y)$  – коэффициент ковариации совокупностей  $X$  и  $Y$ ;

$r_{XY}$  – коэффициент корреляции.

Запишем рассчитанные результаты (табл. 9).

Получим:  $\sigma_{XY} = 0.087825$  и  $r_{XY} = 0.836325$ .

Заметим, что СКО составляет 0.087825 при значении коэффициента корреляции 0.836325, что является приемлемым значением при рассмотре-

нии вопроса подобия двух совокупностей. Следовательно, можно сделать вывод, что имеет место репрезентативная выборка [9, 11].

Для наглядности полученных результатов построим гистограмму распределения вершин для генеральной совокупности и полученной выборки (по горизонтальной оси – вершины с определенной степенью, по вертикальной – количество вершин данной степени), иллюстрирующих полученные ранее результаты (рис. 8):

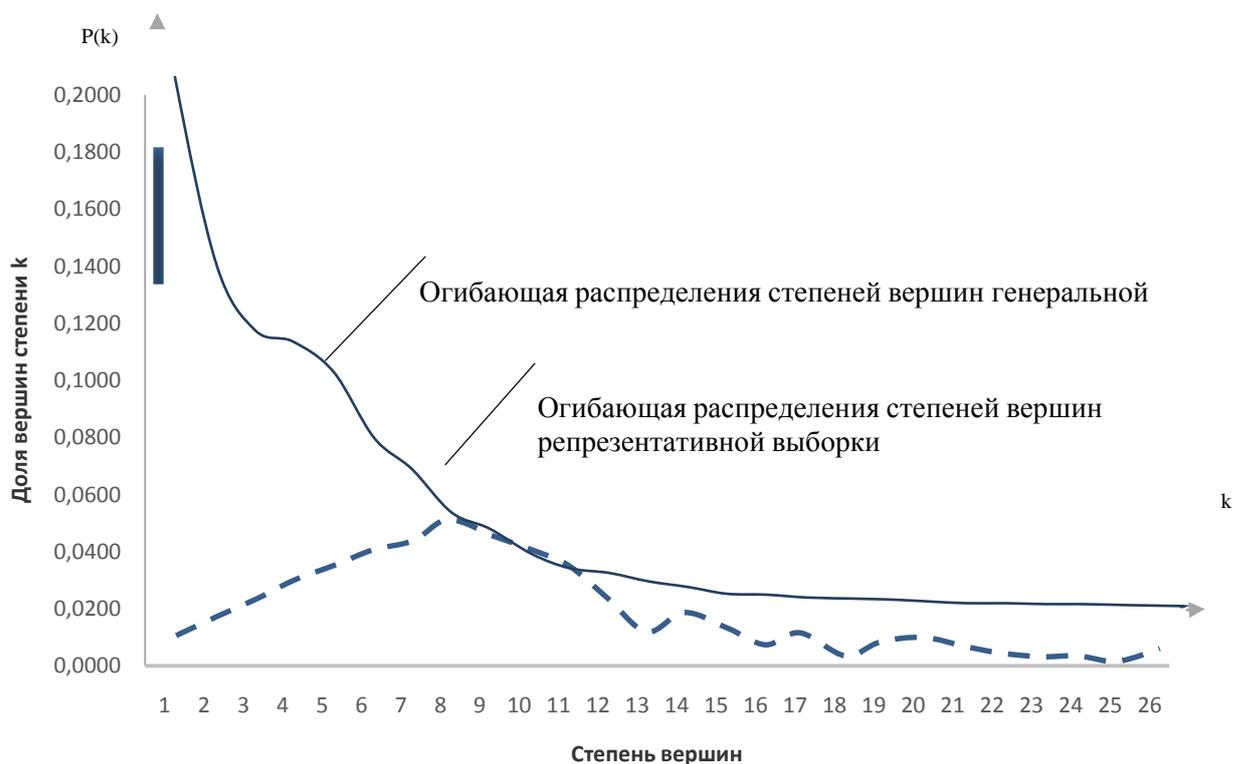


Рис. 8. Распределение количества вершин по степеням

Опираясь на передовые исследования отечественных и зарубежных ученых [1-7], был выполнен анализ исходных статистических данных социальной сети Foursquare, удалось построить структурно-функциональную модель, учитывающую особенности рассматриваемой социальной сети, визуализировать модель усеченной сети и получить структурированные данные описывающие распространение трафика внутри сети.

## Литература:

1. Analytical estimation of the component viability of distribution automated information data system / G.A. Ostapenko, D.G. Plotnicov, O.Y. Makarov, N.M. Tikhomirov, V.G. Yurasov // World Applied Sciences Journal. – 2013. – 25 (3). – P. 416-420.
2. Analytical models of information-psychological impact of social information networks on users / G.A. Ostapenko, L.V. Parinova, V.I. Belonozhkin, I.L. Bataronov, K.V. Simonov // World Applied Sciences Journal. – 2013. – 25 (3). – P. 410-415.
3. Assessment of the system's EPI-resistance under conditions of information epidemic expansion / N.M. Radko, A.G. Ostapenko, S.V. Mashin, O.A. Ostapenko, D.V. Gusev // Biosciences Biotechnology Research Asia. – 2014. – Vol. 11 (3). – P. 1781-1784.
4. Discreet risk-models of the process of the development of virus epidemics in non-uniform networks / V.V. Islamgulova, A.G. Ostapenko, N.M. Radko, R.K. Babadzhanov, O.A. Ostapenko // Journal of Theoretical and Applied Information Technology. – 2016. – P. 306-315.
5. Flood-attacks within the hypertext information transfer protocol: damage assessment and management / A.G. Ostapenko, M.V. Bursa, G.A. Ostapenko, D.O. Butrik // Biosciences Biotechnology Research Asia. – 2014. – Vol. 11 (Spl.End). – P. 173-176.
6. Optimization of expert methods used to analyze information security risk in modern wireless networks / S.A. Ermakov, A.S. Zavorykin, N.S. Kolenbet, A.G. Ostapenko, A.O. Kalashnikov // Life Science Journal. – 2014. – № 11(10s). – P. 511-514.
7. Peak risk assessing the process of information epidemics expansion / N.M. Radko, A.G. Ostapenko, S.V. Mashin, O.A. Ostapenko, A.S. Avdeev // Biosciences Biotechnology Research Asia. – 2014. – Vol. 11. – P. 251-255.

8. Understanding venue popularity in Foursquare / Xuelian Long, Lei Jin, James Joshi // University of Pittsburgh. – 2015. – P. 10-26.
9. Теория вероятностей и математическая статистика / Гмурман В.Е. // Учебное пособие. Высшее образование. - 2007. – 290 с.
10. Социальные сети: модели информационного влияния, управления и противоборства / Губанов Д.А., Новиков Д.А., Чхаратишвили А.Г. Под ред. чл.-корр. РАН Д.А. Новикова // - М.: Издательство физико-математической литературы. – 2010. – 288 с.
11. Статистический анализ данных, моделирование и исследование вероятностных закономерностей. Компьютерный подход: / Лемешко Б.Ю., Лемешко С.Б. // Измерительная техника. – 2008. – 307 с.
12. Средство визуализации данных gephi. - Электрон. Дан. – Режим доступа: <https://gephi.org/>.
13. База данных статистики по социальным сетям - Электрон. Дан. – Режим доступа: <https://github.com/jalbertbowden/foursquare-user-dataset>.
14. Большая Стенфордская Коллекция Сетевых Данных - Электрон. Дан. – Режим доступа: <https://snap.stanford.edu/data>.
15. Анализ контента социальных медиа в эпоху больших данных - Электрон. Дан. – Режим доступа: <https://mebius.io/analysis/social-media-content-analysis>.
16. Социальная сеть Foursquare. - Электрон. Дан. – Режим доступа: <http://foursquare.com>.
17. Статистические данные посещения социальной сети. - Электрон. Дан. – Режим доступа: <http://www.alexa.com>.

Воронежский научно-образовательный центр  
управления информационными рисками

УДК 004.056.57

**СТРУКТУРНО – ФУНКЦИОНАЛЬНАЯ СПЕЦИФИКА  
РАСПРОСТРАНЕНИЯ КОНТЕНТА В СОЦИАЛЬНОЙ СЕТИ  
TOUROUT**

**Д.И. Филиппов, А.В. Паринов**

TourOut.ru — первая и крупнейшая по аудитории в Рунете социальная сеть для туристов и путешественников. Сеть носит по своему наполнению развлекательный и ознакомительный характер и относится к сетям отзывов и обзоров [1].

Регистрация пользователей делится на 2 части – можно зарегистрироваться как обычный пользователь, либо как профессионал в туризме, который должен подтвердить свою специализацию соответствующим документом [8].

Профиль пользователя имеет гибкую настройку, что позволяет выдавать только ту информацию, которая необходима пользователю, скрывая от поиска и просмотра ненужные места и обзоры. Выбирая место своей дислокации, будут предложены различные варианты совместных поездок с людьми, которые расположены в ближайших местах [9].

В настройках приватности есть возможность закрыть свой профиль от незарегистрированных пользователей или пользователей с маленьким рейтингом, что дает возможность защитить свой аккаунт от возможности получения спама или вредоносного контента [15].

Каждый пользователь может оценить обзор, написав комментарий, пожаловаться или поблагодарить автора [1].

В рассматриваемой социальной сети Tourout выделяют 2 основные категории ресурсов: ресурсы коллективного пользования и персональные ресурсы пользователя (рис.1).

Ресурсы коллективного пользования доступны всем пользователям сети и включают в себя отзывы, посты, интерактивную карту и группы. В каждой группе присутствует, как минимум, администратор и модератор, которые могут запретить доступ к группе любому пользователю за нарушение правил поведения в данной группе. В этом случае пользователь получает «бан» и сможет обратно вернуться в группу по прошествии времени, которое было указано в блокировке. В большинстве случаев от 1 дня до 1 месяца [9].

Интерактивная карта представляет собой визуальное представление всех обзоров в данной социальной сети [8].

Каждый отзыв, публикуемый на сайте, проходит проверку у модератора социальной сети. Если в тексте будет замечен спам, большое содержание рекламных продуктов, экстремистские тексты и т.д., отзыв будет отклонен и направлен автору для редактирования, с указанием тех частей текста, где нужно исправить содержание. После исправления и повторной проверки модератором, в случае его одобрения, материал публикуется на сайте и становится доступным для всех участников сети. Каждый пользователь сможет с ним ознакомиться, поставить свою оценку, которая представлена в рейтинговой системе, написать комментарий, отблагодарить автора, поделиться публикацией, пожаловаться на материал, с указанием причины [1].

К персональным ресурсам пользователя относятся ресурсы, к которым имеет доступ только определённый пользователь, остальные участники социальной сети, например, смогут просмотреть профиль пользователя, но изменить информацию в этом профиле они не смогут. Каждый пользо-

ватель настраивает свой профиль, исходя из своих соображений безопасности и конфиденциальности, кто-то указывает максимально возможную информацию о себе, своих интересах, путешествиях, предпочтениях, свои фото и видео материалы, при этом формирует свою ленту максимально точно для отображения, а кто-то не указывает о себе ничего конкретного, только общие пункты, которые не дают полной картины о человеке и о его интересах, в этом случае лента пользователя может быть составлена не точно и содержать материалы, которые будут не интересны участнику сети [9].

Лента конкретного пользователя недоступна никому и является практически уникальным ресурсом для каждого отдельно взятого пользователя, т.к. она формируется на основе настроек профиля, интересов, групп и рекомендаций для конкретного участника социальной сети [1].

Субъекты социальной сети Tourout представлены в виде двух групп: активные и пассивные пользователи (рис. 2) [15].

Активные пользователи являются ценными участниками социальной сети, на которых базируется все основное наполнение сайта и включают в себя такие субъекты, как системный администратор, пользователь, генерирующий контент, модератор и пользователь, распространяющий контент. Системный администратор контролирует правильность работы серверов и устраняет системные ошибки в процессе работы, генерирующий контент пользователь – создает контент или размещает не уникальный контент с других сайтов, модератор – контролирует размещение контента зарегистрированными пользователями в сети и блокирует участников, нарушивших правила социальной сети Tourout, распространяющий контент пользователь – размещает на своей странице уже имеющийся в сети контент, путем копирования или репоста [8].

Пассивные – незарегистрированные пользователи. Также пассивными пользователями являются участники социальной сети, которые зарегистрированы, но не несут пользы – их активные действия минимальны [2].

В сети Tourout представлены 3 главных вида контента: текстовая информация, фото и геолокация (рис. 3). Наличие более 200000 обзоров, говорит о том, что текстовая информация преобладает в данной сети. Каждодневно пользователи пишут и публикуют на сайт обзоры, с возможностью прикрепления фотосета и указания местоположения на интерактивной карте.

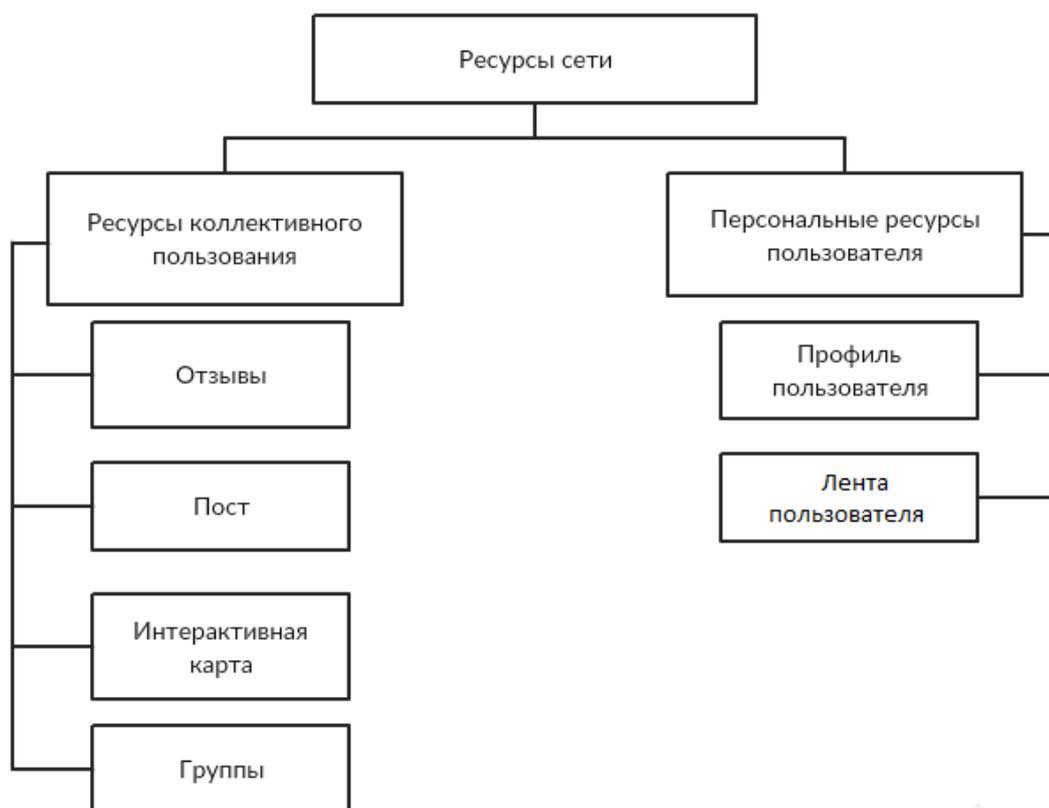


Рис. 1. Классификация объектов сети TourOut



Рис. 2. Классификация субъектов сети Tourout

В силу того, что все публикуемые материалы сначала проходят проверку модератора сайта, чтобы попасть в соответствующий раздел, количество вредоносного контента в данном типе информации минимально. Иначе обстоит дело с комментированием обзоров и отзывов. Комментарии модераторы сразу не проверяют, а вредоносный контент могут заметить спустя некоторое время или при первой жалобе какого-либо пользователя на сообщение. В социальной сети встроена защита от массовой накрутки комментариев. Во-первых, чтобы добавить комментарий пользователь должен доказать, что он «не робот», введя в специальное окно проверяемую информацию, которую сложно распознать компьютерной программе, во-вторых установлены лимиты на добавление комментариев – 10 комментариев в час. В случае, если модератор увидит явную накрутку комментариев со спамным содержанием – пользователь будет заблокирован, а его IP

внесен в «черный список» злоумышленников. Доступ к сайту с данного IP и с данного профиля будет недоступен на время, указанное в блокировке [5,14].

У каждой публикации – отзыва или обзора имеется шкала для оценки. Каждый зарегистрированный пользователь сможет оценить публикацию по 9-ти бальной шкале, выставив соответствующее количество «звезд» [1].

При этом обзор, набравший от 1 до 3 «звезд» имеет низкий рейтинг, от 4 до 6 – средний рейтинг и от 7 до 9 – высокий рейтинг. Данный рейтинг имеет роль при сортировке отзывов и обзоров в новостной ленте – чем больше рейтинг, тем выше обзор находится в списке [8].

Для пользователей, которые впервые на сайте имеются разделы для ознакомления с социальной сетью и ее основными функциями [15].

Все отзывы и обзоры имеют собственный уровень доступа, который позволяет контролировать просмотр и комментирование публикации только лишь определенному кругу пользователей. Существует 3 вида уровня доступа:

1) Доступный всем – открытые публикации для просмотра всем пользователям и комментирования авторизованным пользователям.

2) Доступен только авторизованным пользователям – публикации доступны для просмотра и комментирования только авторизованным пользователям.

3) Доступен определенному кругу лиц – публикации могут просматривать и комментировать пользователи, которые подходят под один из критериев: состоят в определенной группе, имеют высокую репутацию [7].

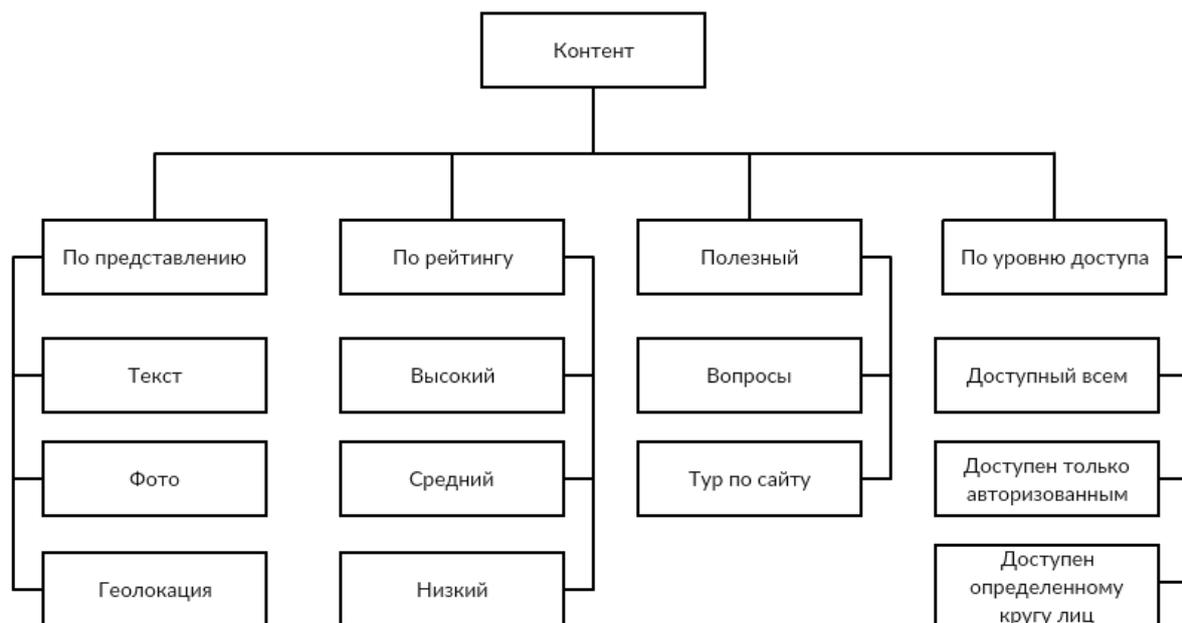


Рис. 3. Классификация разновидности циркулирующего контента в сети  
Tourout

Несмотря на то, что весь пользовательский контент проходит строгую премодерацию, на просторах сети Tourout очень часто встречаются публикации, содержащие в себе вредоносный контент. Пользователи, имеющие большую репутацию на сайте имеют возможность редактировать свои обзоры после публикации (не более 20% содержимого). Этим моментом и пользуются хакеры, которые взламывают аккаунты «премиум-пользователей» и используют их в качестве распространителей вирусного контента, размещая в обзорах ссылки на зараженные сайты, оказывающие деструктивное воздействие на пользователей, перешедших на сайт [8].

Действия пользователей делятся на 3 группы (рис. 4):



Рис. 4. Классификация взаимодействия субъектов социальной сети

1) **Размещение** – действия, выполняемые пользователем, который первоначально создает новый сектор социальной сети. К этой группе относятся такие действия как: создание аккаунта, публикация постов и создание группы. Последние же два пункта невозможно совершить без создания аккаунта. Размещать контент могут только лишь зарегистрированные пользователи.

2) **Ознакомление** – процедура просмотра информации с целью получить новую, интересную и актуальную информацию. К этой группе действий относится чтение и просмотр интерактивной карты. Все эти действия могут совершать совершенно любые пользователи сети, в независимости от их регистрации на сайте.

3) **Реагирование** – активное действие, совершенное пользователем на основе просмотренной публикации. Это может быть лайк, репост

или добавление комментария. Все эти функции, кроме репоста, доступны только зарегистрированным пользователям [15].

Функционал зарегистрированных пользователей намного обширнее, чем незарегистрированных и позволяет контролировать все действия определённого пользователя, ведя журнал совершенных действий у руководства сети [3].

На основе предыдущих классификаций представим структурно-функциональную схему сети Tourout (рис. 5). В сети представлены следующие субъекты: администраторы, модераторы, активные и пассивные пользователи. Администраторы сети предоставляют права модераторам, устанавливают настройки и исправляют неполадки. Модераторы в свою очередь проверяют контент на уникальность, спам и экстремистское содержание, а также контролируют пользователей в группах. Имеют возможность изменить или удалить любую информацию в группах пользователей [4].

Активными пользователями являются участники сети, которые зарегистрированы на сайте и проявляют активность своими действиями, такими как – комментирование, оценивание и распространение контента в другие социальные сети [10].

Пассивные пользователи — это незарегистрированные пользователи или пользователи, которые имеют аккаунт, но не несут ценности в своих действиях, основное действие — это ознакомление с контентом. Ресурсами сети являются: аккаунты пользователей, интерактивная карта, на которой отображаются посещаемые места и города, лента активности пользователей и группы, в которых объединены пользователи по каким-либо интересам. Группы являются отличительной особенностью данной социальной сети по сравнению с другими сетями, относящимися к категории обзоров [1].

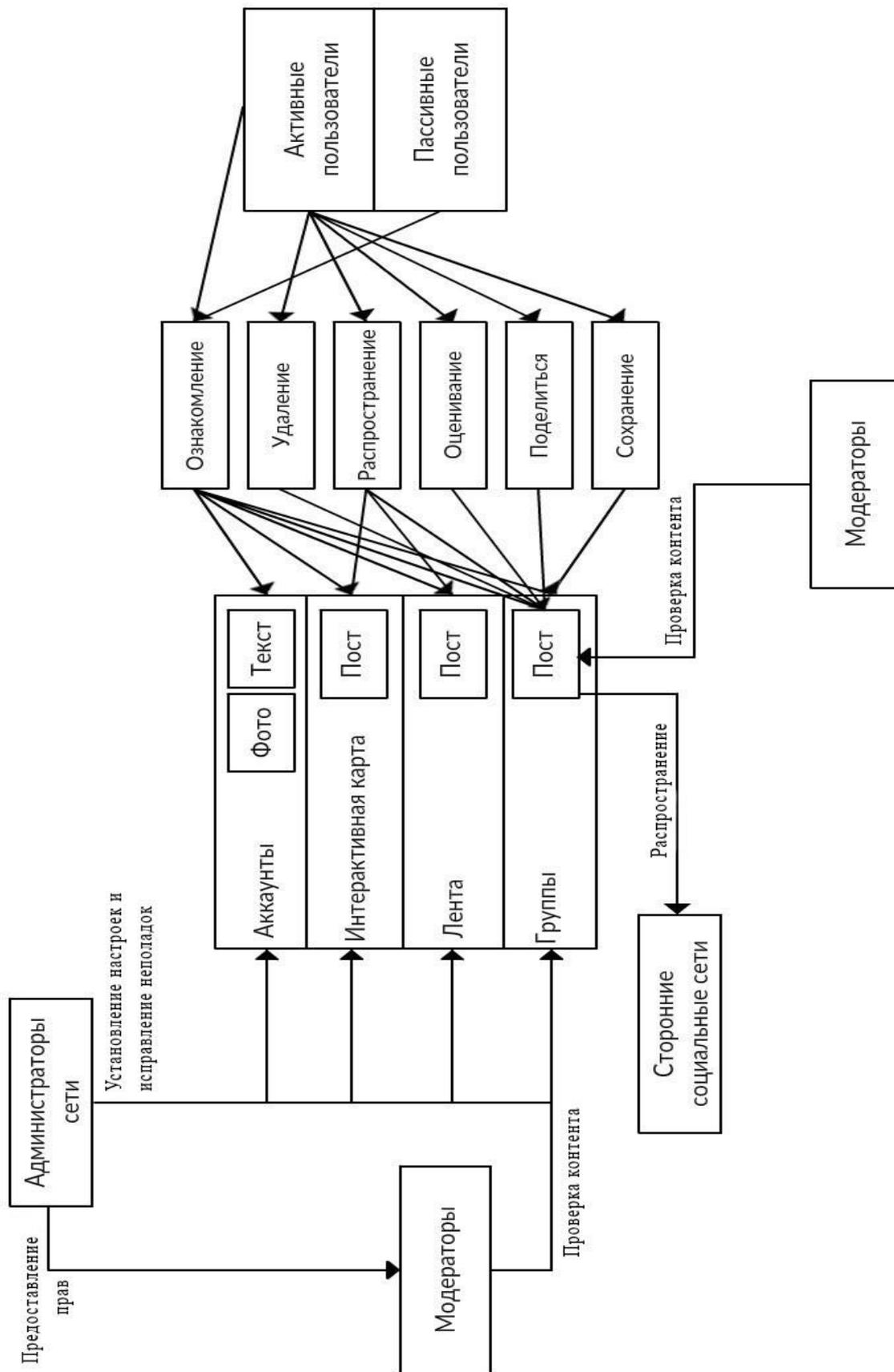


Рис. 5. Структурно-функциональная схема сети Tourout

Объединения пользователей в определённые группы помогает злоумышленникам выделить из всех пользователей только нужную им категорию людей и провести узконаправленную информационную атаку. Основной контент, представленный в данной социальной сети – это текст, фото и геолокационная метка [1].

Построим звездную матрицу взвешенной связности исходной сети (табл. 1):

Таблица 1 – Звездная матрица социальной сети Tourout

Номер узла	1	2	3	...	815	816	...	30501	30502
1	0	5	2	...	0	0	...	0	0
2	0	0	0	...	0	0	...	0	0
3	0	0	0	...	3	0	...	0	0
4	0	3	0	...	0	0	...	0	0
...	...	...	...	...	...	...	...	...	...
815	0	0	4	...	0	0	...	0	0
816	0	0	0	...	0	0	...	0	0
...	...	...	...	...	...	...	...	...	...
30501	0	0	0	...	0	0	...	0	0
30502	0	0	0	...	0	0	...	0	0

Для последующей репрезентативной выборки из мультиразмерной сети необходимо знать удельный вес ее вершин и дуг. Для этого предлагается нормировка их весов по сумме весов всех дуг (1) сети

$$\sum_{i \neq j} \delta(a_{ij}), \quad (1)$$

т.е. – по суммарному трафику сети. Тогда нормированная величина (2), полученная в данном случае

$$\delta(\bar{a}_{ij}) = \delta(a_{ij}) / \sum_{i \neq j} \delta(a_{ij}), \quad (2)$$

будет показывать удельный вес трафика в дуге  $a_{ij}$  к суммарному трафику сети. Она и будет характеризовать степень ее взвешенной (по трафику) центральности.

Следует заметить, что суммарный трафик сети не разделяет входящие и исходящие дуги. Поэтому для определения взвешенной центральности вершины  $x_s$  можно использовать сумму (3)

$$\sum_i \delta(a_{si}) + \sum_j \delta(a_{js}), \quad (3)$$

которую далее следует пронормировать по суммарному трафику сети. В результате получим нормированную величину (4)

$$\delta(\bar{x}_s) = [\sum_i \delta(a_{si}) + \sum_j \delta(a_{js})] / \sum_{i \neq j} \delta(a_{ij}), \quad (4)$$

которая будет характеризовать удельный вес трафика, проходящего через вершину  $x_s$ , по отношению ко всему трафику сети. Это и будет характеризовать степень взвешенной (по трафику) центральности данной вершины.

Следуя вышеизложенному, можно получить квадратную матрицу взвешенной центральности элементов исходной сети (табл. 2).

Исходя из статистических данных, характеризующих количество узлов и ребер изобразим граф, полученный на основе исходных данных, в котором вершины являются пользователями, а взвешенные дуги обозначают количество отзывов и обзоров в месяц (рис. 6). На графе явно прослеживается кластерность. Следствием является высокий коэффициент кластеризации (табл. 3) [12,16].

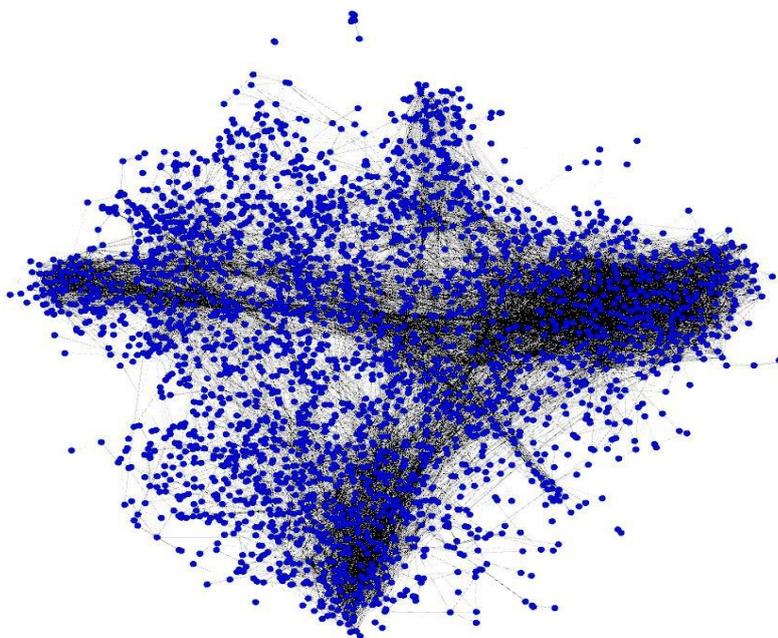


Рис. 6. Визуализация исходного графа

Таблица 2 – Матрица взвешенной центральности для сервиса Tourout

Номер узла	1	2	...	815	⋮	30501	30502
1	0,0007574	0,0019823	...	0	⋮	0	0
2	0	0,0005957	...	0	⋮	0	0
...	⋮	⋮	⋮	⋮	⋮	0	0
814	0,005128	0	...	0	⋮	0	0
815	0	0	...	0,00012307	⋮	0	0
...	...	...	...	...	⋮	0	0
30501	0	0	0	0	...	0,0000096	0
30502	0	0	0	0	...	0	0,00018

Таблица 3 – Статистические данные социальной сети Tourout

Характеристические данные	Показатели
Количество узлов	30502
Количество ребер	560383
Диаметр	18
Средняя длина пути	2,85
Коэффициент кластеризации (%)	1,004
Средний диаметр сети	3,7291

Рассмотрим диагональные элементы матрицы взвешенной центральности, полученной выше. Далее произведем сортировку вершин, по убыванию оценки взвешенной центральности [14].

Социальная сеть Tourout состоит из нескольких тысяч вершин, что делает трудоемким полный анализ сети. Преобразуем сеть в выборку, которая по своей структуре будет отражать основные свойства нашей сети [1].

Для получения репрезентативной выборки воспользуемся алгоритмом, реализованным совместно с партнерами по комплексной работе. За допустимую погрешность примем не более 5% трафика для усеченной модели сети.

Алгоритм получения репрезентативной выборки:

1. Вычленив из исходной матрицы взвешенной центральности элементов сети ее диагонали. Оставшаяся усеченная матрица с обнуленными диагональными элементами будет использована на шаге 4.

2. Отсортировать полученную матрицу взвешенной центральности по убыванию значений степени взвешенной центральности вершин  $\delta(\bar{x}_s)$ , расположенных на главной диагонали.

3. Найти минимальную сумму значений  $\delta(\bar{x}_s)$ , такую что  $\sum \delta(\bar{x}_s) > 0.95$  и исключить остальные значения.

4. Удалить из усеченной матрицы все элементы, относящиеся к исключенным на предыдущем шаге путём обнуления столбцов и строк.

Таблица 4 – Степени взвешенной центральности вершин

Номер узла	Степень
1	0,0007574
2	0,0005957
3	0,00075277
...	...
30501	0,000096
30502	0,000018

Просуммируем значения для получения необходимого результата (точности модели). В нашем случае до значения 0.95, так как для нас допустима 5% потеря трафика. Этим критерием и ограничится репрезентативная выборка:

$$\begin{aligned} \delta(\bar{x}_g) &= 0,0291092 + 0,0287001 + 0,0270092 + \dots + 0,0090480 \\ &\quad + 0,0088901 + \dots + 0,0000075049 + 0,0000075001 \\ &= 0,9507. \end{aligned}$$

При данном суммировании получили выборку из 341 элемента.

Далее обнулим столбцы и строки, на пересечении которых стояла исключенная вершина.

На основе данного усечения построим проранжированную усеченную матрицу (табл. 5):

Таблица 5 – Проранжированная усеченная матрица социальной сети Tourout

Номер узла	1701	967	...	2008	...	30020	29808
1701	0	0	...	0	...	0	0
967	0,00015787	0	...	0	...	0	0
...	⋮	⋮	↘	⋮	⋮	⋮	⋮
101	0,00010877	0	...	0	...	0	0
2008	0	0		0		0,0002688	0
...	⋮	⋮	⋮	⋮	↘	⋮	⋮
30020	0	0	...	0	...	0	0
29808	0	0	...	0	...	0	0

На основе данной матрицы, с помощью программного обеспечения Gephi [15] построим усеченный граф социальной сети (рис. 7):

Стоит отметить, что у данной социальной сети прослеживается кластерность. Это напрямую зависит от групп пользователей, которые интересуются определенной страной или местом, тем самым для удобного обмена информацией вступают в соответствующие группы. Обратим внимание,

что кластерность присущая исходному графу сохранилась в выборке, что подтверждает верность расчётов по выбранному алгоритму.

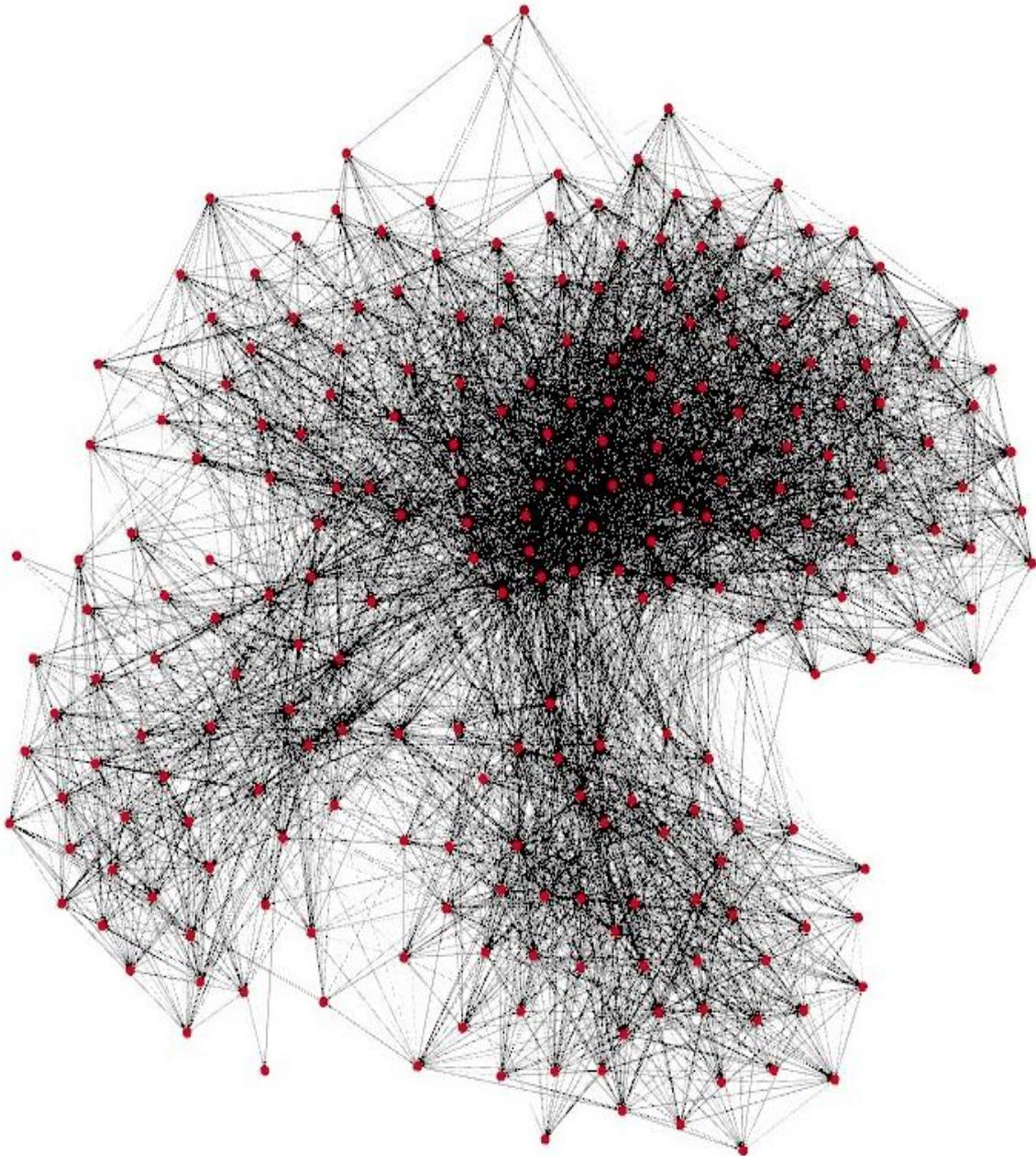


Рис. 7. Граф, построенный на основе полученной выборки

Для моделирования процессов информационной диффузии в сети Tourout построим модифицированную матрицу смежности социальной сети Tourout. Данная матрица строится на основе двух правил:

1. Недиagonальные элементы равны  $\pm 1$ , так как интересуется лишь факт смежности вершин истоков (+1) и стоков (-1).

2. Диагональные элементы имеют вес  $k$ , равный сумме элементов (+1) столбца.

Таблица 6 – Модифицированная матрица смежности социальной сети Tourout

Номер узла	1701	967	...	101	2008	...	30020	29808
1701	3	0	...	0	0	...	0	0
967	0	2	...	0	0	...	0	0
...	⋮	⋮	↘	⋮	⋮	⋮	⋮	⋮
101	0	0	...	8	0	...	0	0
2008	0	0	...	0	5	0	0	0
...	⋮	⋮	⋮	⋮	⋮	↘	⋮	⋮
30020	0	0	...	0	0	...	0	0
29808	0	0	...	0	0	...	0	1

После применения алгоритма преобразования исходных данных сети и нахождения репрезентативной выборки была получена визуальная модель социальной сети Tourout и соответствующие матрицы для дальнейшего анализа социальной сети (рис. 7) [5, 17].

Построим таблицу распределения генеральной совокупности, представляющие варианты данной совокупности (степени вершин) и их частоту (табл. 7).

Таблица 7 – Генеральная совокупность

$i$	1	2	3	...	33	34	35	...	122	123	124
$x_i$	1	2	3	...	12	18	20	...	201	213	237
$n_i$	1520	1152	734	...	10	14	12	...	2	2	2

Проведем аналогичные расчёты для выборочной совокупности (табл. 8).

Таблица 8 – Выборочная совокупность

i	1	2	3	...	10	11	12	...	22	23	24
$x_i$	2	3	4	...	21	22	23	...	107	111	119
$n_i$	2	2	2	...	10	10	24	...	2	2	2

Параметры графа выборки представим в виде таблицы (табл. 9):

Таблица 9 – Параметры графа выборки

Параметры графа	Значения для выборки
Количество узлов	341
Количество различных степеней	24
Диаметр графа	<b>21</b>
Средняя длина пути	8.18

Проведем сравнение распределений количества вершин по степеням для исходного графа и для графа выборки (рис. 8).

Для доказательства генеральной и выборочной совокупности воспользуемся критерием Пирсона.

$$\chi^2_{\text{набл}} = \sum_{i=1}^k \frac{(n - np_i)^2}{np_i}, \quad (5)$$

где  $n$  – объем выборки;

$p_i$  - вероятность попадания случайной величины  $X$ .

Следуя представленной формуле (5), рассчитаем значение  $\chi^2$ . В данном случае  $\chi^2=24.3$ . Следовательно,  $n = 24$  (выборка) – 1 (количество параметров в плотности степенного распределения) = 23 (число степеней свободы) [14].

Выберем уровень значимости  $\gamma=0,05$ , тогда  $p=1-\gamma=0,95$ . Табличное значение  $\chi^2(23)$  равно 35.1. Получается, что табличное значение  $\chi^2$  больше

вычисленного. Следовательно, гипотеза верна, и приведенное степенное распределение согласуется с полученными результатами.

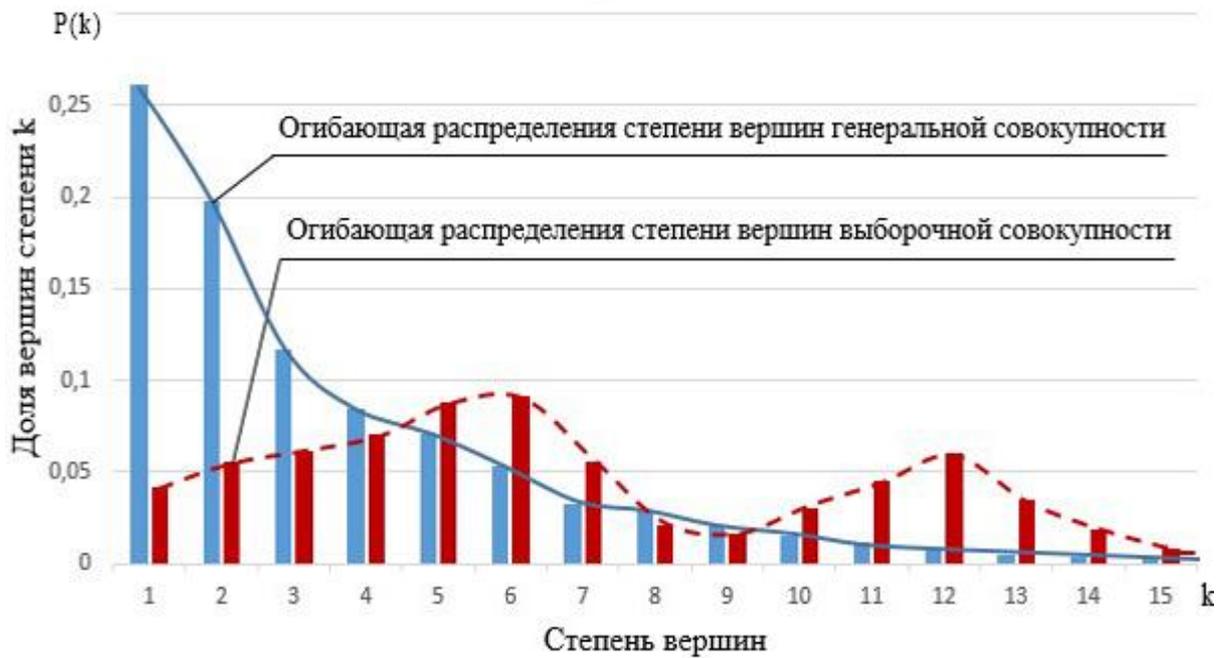


Рис. 8. Распределение количества вершин по степеням

Для доказательства репрезентативности выборки вычислим параметры, описывающие связность значений исходного распределения долей и значений выборочной совокупности. Рассчитаем значения среднеквадратичного отклонения (СКО) по формуле (6) значений двух совокупностей и коэффициент корреляции (6-9) [13].

$$\sigma_{XY} = \sqrt{\sum_{i=1}^n (X_i - Y_i)^2}, \quad (6)$$

$$cov(X, Y) = \frac{1}{n} \sum_{t=1}^n (X_t - \bar{X})(Y_t - \bar{Y}), \quad (7)$$

$$\sigma_X = \sqrt{\frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^2}, \quad (8)$$

$$r_{XY} = \frac{cov(X, Y)}{\sigma_X \sigma_Y}, \quad (9)$$

где:  $n$  – общее число вершин;

$X_i$  –  $i$ -ый элемент выборки;

$\bar{X}$  – среднее арифметическое выборки;

$\sigma_X, \sigma_Y$  – среднеквадратическое отклонение для совокупностей  $X$  и  $Y$  соответственно;

$\sigma_{XY}$  – среднеквадратическое отклонение значений выборочной совокупности от генеральной;

$cov(X, Y)$  – коэффициент ковариации совокупностей  $X$  и  $Y$ ;

$r_{XY}$  – коэффициент корреляции.

Запишем рассчитанные результаты (табл. 10).

Таблица 10 - Рассчитанные значения сравнения совокупностей

Коэффициент	Значение
СКО	0,0720914
Коэффициент корреляции	0,7879281

Значения СКО и коэффициент корреляции являются допустимыми и, следовательно, имеет место репрезентативная выборка.

Таким образом, опираясь на передовые исследования отечественных и зарубежных ученых [2-8, 11], был проведен анализ и систематизация исходных статистических данных социальной сети Tourout, построена структурно-функциональная модель и была визуализирована модель усеченной

сети, также получены структурированные данные описывающие распространение трафика внутри сети.

#### Литература:

1. About leading russian travel social network & tourists web portal, TourOut.ru - Электрон. Дан. – Режим доступа: <http://tourout.ru/text/about.html>
2. Analytical estimation of the component viability of distribution automated information data system / G.A. Ostapenko, D.G. Plotnicov, O.Y Makarov, N.M. Tikhomirov, V.G. Yurasov // World Applied Sciences Journal. – 2013. – 25 (3). – P. 416-420.
3. Analytical models of information-psychological impact of social information networks on users / G.A. Ostapenko, L.V. Parinova, V.I. Belonozhkin, I.L. Bataronov, K.V. Simonov // World Applied Sciences Journal. – 2013. – 25 (3). – P. 410-415.
4. Assessment of the system's EPI-resistance under conditions of information epidemic expansion / N.M. Radko, A.G. Ostapenko, S.V. Mashin, O.A. Ostapenko, D.V. Gusev // Biosciences Biotechnology Research Asia. – 2014. – Vol. 11 (3). – P. 1781-1784.
5. Discreet risk-models of the process of the development of virus epidemics in non-uniform networks / V.V. Islamgulova, A.G. Ostapenko, N.M. Radko, R.K. Babadzhanov, O.A. Ostapenko // Journal of Theoretical and Applied Information Technology. – 2016. – P. 306-315.
6. Flood-attacks within the hypertext information transfer protocol: damage assessment and management / A.G. Ostapenko, M.V. Bursa, G.A. Ostapenko, D.O. Butrik // Biosciences Biotechnology Research Asia. – 2014. – Vol. 11 (Spl.End). – P. 173-176.

7. Optimization of expert methods used to analyze information security risk in modern wireless networks / S.A. Ermakov, A.S. Zavorykin, N.S. Kolenbet, A.G. Ostapenko, A.O Kalashnikov // Life Science Journal. – 2014. – № 11(10s). – P. 511-514.
8. Peak risk assessing the process of information epidemics expansion / N.M. Radko, A.G. Ostapenko, S.V. Mashin, O.A. Ostapenko, A.S. Avdeev // Biosciences Biotechnology Research Asia. – 2014. – Vol. 11. – P. 251-255.
9. Sampling from large graphs / Leskovec J., Faloutsos C. // Science Journal. - 2006. – 631p.
10. Social Network Analysis in Telecommunications / Pinheiro A. // Science Journal. — 2011. — Vol. 4. — P. 100-102.
11. Алферов А.Г. Устойчивость инфокоммуникационных систем в условиях информационного конфликта [Текст]/ А.Г. Алферов, И.О. Толстых, Н.Н. Толстых, О.В. Поздышева, А.И. Мордовин // Информация и безопасность. – 2014. – Т. 17. – № 4. – С. 558-567.
12. Дешина А.Е. Атака на серверы резервного копирования: оценка жизнестойкости [Текст]/ А.Е. Дешина, Д.Г. Плотников, Д.С. Каребин, Я.М. Каценко, Э.Д. Поликарпов // Управление информационными рисками и обеспечение безопасности инфокоммуникационных систем: Сб. науч. тр.; под ред. чл.-корр. РАН В.И. Борисова. – Воронеж: Издательство «Научная книга». – 2015. – Т. 8. – № 2. – С. 109-119.
13. Плотников Д.Г. Диапазонный риск-анализ ИТКС на основе логлогистического закона распределения ущерба [Текст] / Д.Г. Плотников, Д.Б. Борисов, О.Ю. Макаров // Информация и безопасность. – 2012. – Т. 15. – № 1. – С. 193-140.
14. Социальные сети: модели информационного влияния, управления и противоборства /Губанов Д.А., Новиков Д.А., Чхаратишвили А.Г. Под

ред. чл.-корр. РАН Д.А. Новикова // - М.: Издательство физико-математической литературы. – 2010. – 288 с.

15. Средство визуализации данных. - Электрон. Дан. – Режим доступа: <https://gephi.org/>

16. Теория вероятностей и математическая статистика / Гмурман В.Е. // Учебное пособие. Высшее образование. - 2007. – 290 с.

17. Шварцкопф Е.А. Моделирование эпидемического процесса заражения пользователей безмасштабной сети с учетом ее топологии [Текст]/ Е.А. Шварцкопф, Ю.Н. Гузев, И.Л. Батаронов, В.И. Белоножкин, К.А. Разинкин // Информация и безопасность. – 2015. – Т. 18. – № 4. – С. 520-523.

Воронежский научно-образовательный центр  
управления информационными рисками

УДК 004.096.57

**НАУЧНО-МЕТОДИЧЕСКОЕ ОБОБЩЕНИЕ ПРОЦЕДУР  
МОДЕЛИРОВАНИЯ ИНФОРМАЦИОННОЙ ДИФФУЗИИ В  
СОЦИАЛЬНЫХ СЕТЯХ**

**А.В. Паринов, В.В. Морковина, А.А. Остапенко**

В рассматриваемом контексте можно выделить три основные процедуры:

- преобразования исходных данных об исследуемой социальной сети, которые имеются в Интернете;
- репрезентативной выборки из анализируемой соцсети с целью снижения размерности моделируемого объекта;
- непосредственного моделирования процессов диффузии контента на основе данных, полученных в результате выполнения вышеперечисленных процедур.

Рассмотрим обозначенные процедуры подробнее.

**Алгоритм преобразования исходных данных сети**

Фактически стоит задача создания базы данных соцсетей (желательно самой широкой) для их топологического описания.

В настоящее время взвешенные данные инциденции их вершин и дуг имеются в Интернете (см. соответствующие сайты) для всех разновидностей соцсетей. Формат этих данных соответствует **трехместному предикату**

$$\Gamma(x_i, a_{ij}, x_j) \Leftrightarrow \Gamma(i, \delta(a_{ij}), j),$$

где  $i$  и  $j$  – номера вершин  $x_i$  и  $x_j$  в сети;

$\delta(a_{ij})$  - вес дуги  $a_{ij}$ , связывающей  $x_i$  и  $x_j$  и направленной от  $i$  к  $j$ .

Авторы такого формата данных в той или иной степени подразумевают под весом дуги **динамический ресурс**, т.е. передачу определенного объема  $V$  и ценности  $C$  наполнителя сети в единицу времени, т.е.

$$\delta(a_{ij}) = \frac{\partial[cv]}{\partial t} = \langle C \rangle V',$$

где с усредненной ценностью  $\langle C \rangle$  они связывают популярность и др. параметры пользователей  $i$  и  $j$ , а  $V'$  относят к интенсивности обмена контентом.

Такой формат позволяет построить **звездную матрицу**, элементы строки которой соответствуют дугам, входящим в данную вершину, а элементы столбца – дугам, исходящим из вершины. Матрица очевидно будет квадратная, размерности, равной количеству вершин в сети, и ее диагональные элементы будут равны нулю.

Отсюда:

**Шаг 1.** В соответствии с вышеизложенным, формирование звездной матрицы взвешенной связности исходной сети.

**Шаг 2.** Определение **диагональной матрицы степени вершин сети**. Дело в том, что для последующих процедур моделирования процессов диффузии контента нужно для каждой вершины знать количество исходящих из нее дуг, которое назовем ее степенью исхода. Данную матрицу надо хранить отдельным файлом для последующих операций.

**Шаг 3.** Построение **квадратной матрицы взвешенной центральности элементов сети**. Дело в том, что для последующей репрезентативной выборки из мультиразмерной сети необходимо знать удельный вес ее вершин и дуг. Для этого предлагается нормировка их весов по сумме весов всех дуг сети

$$\sum_{\substack{i,j \\ i \neq j}} \delta(a_{ij}),$$

т.е. – по суммарному трафику сети. Тогда нормированная величина, полученная в данном случае

$$\delta(\bar{a}_{ij}) = \delta(a_{ij}) / \sum_{\substack{i,j \\ i \neq j}} \delta(a_{ij}),$$

будет показывать удельный вес трафика в дуге  $a_{ij}$  к суммарному трафику сети. Она и будет характеризовать степень ее взвешенной (по трафику) центральности.

Следует заметить, что суммарный трафик сети не разделяет входящие и исходящие дуги. Поэтому для определения взвешенной центральности вершины можно использовать сумму

$$\sum_i \delta(a_{si}) + \sum_j \delta(a_{js}),$$

которую далее следует пронормировать по суммарному трафику сети. В результате получим нормированную величину

$$\delta(\bar{x}_s) = \left[ \sum_i \delta(a_{si}) + \sum_j \delta(a_{js}) \right] / \sum_{\substack{i,j \\ i \neq j}} \delta(a_{ij}),$$

которая будет характеризовать удельный вес трафика, проходящего через вершину  $x_s$ , по отношению ко всему трафику сети. Это и будет ха-

рактизовать степень взвешенной (по трафику) центральности данной вершины.

Следуя вышеизложенному, можно получить квадратную матрицу взвешенной центральности элементов исходной сети.

**Шаг 4.** Построение диагональной матрицы удельного баланса трафика в вершинах сети. Здесь уже учитывается не сумма и разность исходящего и входящего для вершины трафиков.

$$\delta(\bar{B}_s) = \left[ \sum_i \delta(a_{si}) - \sum_j \delta(a_{js}) \right] / \sum_{\substack{i,j \\ i \neq j}} \delta(a_{ij}),$$

Полученная удельная величина будет характеризовать роль вершины (субъекта) сети:

- при  $\bar{B}_s > 0$  – это генератор контента;
- при  $\bar{B}_s < 0$  - это аккумулятор контента.

Чем больше модуль  $\bar{B}_s$ , тем в большей степени проявляется его ролевая функция. В этом и состоит практическое применение диагональной матрицы удельного баланса трафика в вершине сети.

Описанные шаги по формированию базы данных соцсети следует повторять для каждой ее разновидности, предоставив исследователю возможность наиболее полного анализа класса социальных информационных сетей. При этом, следует помнить, что соцсети постоянно развиваются, и созданная база данных требует своей регулярной актуализации. Если вышеописанные процедуры достаточно успешно автоматизированы, то такая работа не составит особого труда.

Существует и еще одна особенность удельных оценок. Чем больше сеть, тем меньше (по модулю) значения в матрице взвешенной централь-

ности. Это будет создавать некоторые неудобства, но они преодолимы за счет степенно представления чисел

$$\delta(\bar{a}_{ij}) = rE(-d) \& \delta(\bar{x}_s) = qE(-b),$$

где  $d$  – постоянно для записи степени центральности всех дуг и соизмеримо с их количеством в сети;

$b$  – постоянно для записи степени центральности всех вершин и соизмеримо с их количеством в сети. Разрядность  $r$  и  $q$  будет определяться ожидаемым объемом последующие репрезентативной выборки.

### **Процедура репрезентативной выборки из сети**

Мультиразрядность социальных сетей не оставляет надежд для непосредственного использования широкому кругу исследователей тех данных, которые он может найти в Интернете относительно связности элементов сети. Огромные вычислительные затраты делают невозможным использование этих данных в рамках персонального компьютера и, тем более, смартфона.

Поэтому необходима репрезентативная выборка, которая имеет значительно меньшую размерность, но с заданной точностью сохраняет свойства исходной сети.

Тривиальный подход заключается в отборе вершин по их степени, т.е. количеству инцидентным им дуг. Однако, такой критерий явно не учитывает тот факт, что каждая дуга имеет свой вес, т.е. отсутствует взвешенная оценка с точки зрения трафика.

Поэтому уместно использование квадратной матрицы взвешенной центральности элементов сети, а точнее – ее диагонали, где определена степень взвешенной центральности каждой вершины.

Такой подход наиболее объективен, ибо учитывает трафик – основу функционирования любой сети. Отсюда:

**Шаг 1.** Вычленение из квадратной матрицы взвешенной центральности элементов сети ее диагонали, в которой будут сосредоточены только степени взвешенной центральности ее вершин. Оставшаяся усеченная матрица с обнуленными диагональными элементами будет использована в дальнейших процедурах.

**Шаг 2.** Пересортировка последовательности расположения вершин в диагонали. Они должны быть расположены в порядке убывания степени их взвешенной центральности. После такого ранжирования следует последовательно (первая + вторая + третья + ...) суммировать эти степени до тех пор, пока мы не получим требуемую точность модели. Так, если для нас допустима 5% потеря трафика, полученная сумма должна быть не меньше 0,95. Этим критерием и ограничится репрезентативная (с точки зрения трафика) выборка.

**Шаг 3.** Предстоит удалить из усеченной матрицы все ее элементы, относящиеся к исключенным выборкой вершинам. Это делается путем обнуления ее столбцов и строк, на пересечении которых стояла исключенная вершина. Полученная в результате матрица (с обнуленными также диагональными элементами) будет полностью отражать связность компонентов выборки.

**Шаг 4.** Визуализация выборки. Для этого можно пренебречь ориентированным характером дуг и заменить их на неориентированные ребра с единичным весом (на этом этапе нас интересует факт связи). Визуализация может быть реализована на плоскости, но возможен и трехмерный вариант, когда по оси Z откладывается степень центральности вершины. Целью визуализации является проверка выборки на отсутствие несвязных кластеров. Если таковые появились, то надо продолжить добавление в выборку

ку вершин, дабы устранить нарушение целостности сети, появившееся при выборке.

**Шаг 5.** Формирование итоговой матрицы в виде, удобном для последующего моделирования информационной диффузии. Представляется, что она должна выглядеть следующим образом:

- недиагональные элементы равны  $\pm 1$ , так как интересует лишь факт смежности вершин истоков (+1) и стоков (-1). В некоторой степени это возврат к формату матрицы смежности;

- диагональные элементы имеют вес  $k$ , равный сумме элементов (+1) столбца. Отсюда определяется ожидаемое количество вирусированных контентом элементов.

Назовем такую таблицу **модифицированной матрицей смежности**. При этом надо понимать, что это весьма упрощенное представление и, при необходимости учета заметности контента на фоне другого трафика дуги, нужно учитывать ее трафик. В этом случае матрица примет несколько иной вид.

### **Процедура моделирования диффузии контента в сети**

Это алгоритм имеет особую важность, ибо от его адекватности зависят все итоговые риск-оценки. Его основой служит модифицированная матрица смежности.

В первом приближении процесс инфицирования  $k$ -смежных вершин из одного источника может быть описан биномиальным распределением, матожиданием которого является  $[kp]$ , где  $p$  - вероятность единичного заражения, оператор  $[•]$  - вычисляет целую часть произведения, а  $k$  – та самая степень вершины источника инфекции.

Разумеется, это достаточно упрощенный подход, но в первом приближении имеет право на использование.

**Шаг 1.** Задание узлов, из которых планируется запустить контент. Здесь исследователь должен исходить из того, что вероятность  $p$  зачастую весьма невелика, и надо выбирать стартовую вершину с популярностью (степенью), соизмеримой с величиной  $\left(\frac{1}{p}\right)$ . Распараллеливание процесса инфицирования также увеличивает шансы успеха. Но исследователь этот параметр должен иметь возможность задать произвольно по собственному усмотрению.

**Шаг 2.** Задание вероятности единичного инфицирования исследователь может осуществить «с потолка». Однако лучше опираться на некоторый опыт. Например, он может провести эксперимент со своими «друзьями», предложив им оценить сгенерированный им контент. Причем, их множество должно быть достаточно представительным, ввиду малости  $p$ . Важно также понимать, что  $p$  включает в себя возможности: прочтения (заметность), восприятия (осмысления), последующей реакции, иммунитета и др. относительно предлагаемой пользователю информации. То есть конструкция  $p$  может быть сложной, и программные модули, реализующие данный шаг алгоритма, должны предлагать исследователю определенное разнообразие методик вычисления  $p$ .

**Шаг 3.** Исходя из матожидания, следует смоделировать первичный эффект (для каждого источника) заражения. Здесь особую важность имеет высокая степень случайности выбора  $[kp]$  вершин согласно их смежности с первичным источником инфекции. Соответствующий программный модуль это должен четко обеспечивать, тем более, что он будет применяться и при всех последующих (вторичных, третичных и др.) эффектах. Этот шаг заканчивается выбором инфицированных вершин, которые становятся вторичными источниками инфекции.

**Шаг 4.** Рассматривается вторичный эффект, и снова определяется новое множество инфицированных вершин. Это происходит также путем случайного выбора. Уместно заметить, что  $p$  может меняться по ходу процесса в зависимости от: степени вершины, соотношения общего количества уже инфицированных вершин и еще неинфицированных. В идеале программная реализация алгоритма должна учитывать это и предлагать исследователю соответствующие поправки по ходу моделирования.

Описанный шаг будет повторяться вплоть до окончания вычислительного эксперимента, и этот цикл исследователь может задать и/или остановить, когда ему захочется.

**Шаг 5.** Вычисление риска, для реализации которого следует обратиться к диагональной матрице степени взвешенности центральности вершин сети. Величина риска, которая автоматически должна вычисляться на каждом шаге эпидемического процесса, равна сумме вышеупомянутых степеней. По мере развития эпидемии его величина будет расти и приближаться к единице. При этом должна быть предусмотрена соответствующая визуализация динамики этого параметра.

Программная реализация алгоритма должна предусматривать возможность коррекции процесса ввиду снижения актуальности, привлекательности и других параметров контента, что не исключено в ходе вычислительного эксперимента.

В целом рассматриваемый алгоритм зависит от многих факторов, которые хотелось бы учесть, и поэтому его разработка является особо творческой задачей, в решении которой для начала можно опираться на вышеперечисленные шаги, методика реализации которых, разумеется, требует своего развития.

Формализация процедур осуществлена на основе научно-методического обеспечения, разработанного в работах [1-7].

## Литература:

1. Analytical estimation of the component viability of distribution automated information data system / G.A. Ostapenko, D.G. Plotnicov, O.Y. Makarov, N.M. Tikhomirov, V.G. Yurasov // World Applied Sciences Journal. – 2013. – 25 (3). – P. 416-420.
2. Analytical models of information-psychological impact of social information networks on users / G.A. Ostapenko, L.V. Parinova, V.I. Belonozhkin, I.L. Bataronov, K.V. Simonov // World Applied Sciences Journal. – 2013. – 25 (3). – P. 410-415.
3. Optimization of expert methods used to analyze information security risk in modern wireless networks / S.A. Ermakov, A.S. Zavorykin, N.S. Kolenbet, A.G. Ostapenko, A.O. Kalashnikov // Life Science Journal. – 2014. – № 11(10s). – P. 511-514.
4. Assessment of the system's EPI-resistance under conditions of information epidemic expansion / N.M. Radko, A.G. Ostapenko, S.V. Mashin, O.A. Ostapenko, D.V. Gusev // Biosciences Biotechnology Research Asia. – 2014. – Vol. 11 (3). – P. 1781-1784.
5. Peak risk assessing the process of information epidemics expansion / N.M. Radko, A.G. Ostapenko, S.V. Mashin, O.A. Ostapenko, A.S. Avdeev // Bio-sciences Biotechnology Research Asia. – 2014. – Vol. 11 (Spl.End). – P. 251-255.
6. Flood-attacks within the hypertext information transfer protocol: damage assessment and management / A.G. Ostapenko, M.V. Bursa, G.A. Ostapenko, D.O. Butrik // Biosciences Biotechnology Research Asia. – 2014. – Vol. 11 (Spl.End). – P. 173-176.

7. Discreet risk-models of the process of the development of virus epidemics in non-uniform networks / V.V. Islamgulova, A.G. Ostapenko,,N.M. Rad-ko, R.K. Babadzhanov, O.A. Ostapenko // Journal of Theoretical and Applied In-formation Technology. – 2016. – P. 306-315.

Воронежский научно-образовательный центр  
управления информационными рисками

Научное издание

**УПРАВЛЕНИЕ ИНФОРМАЦИОННЫМИ РИСКАМИ  
И ОБЕСПЕЧЕНИЕ БЕЗОПАСНОСТИ  
ИНФОКОММУНИКАЦИОННЫХ СИСТЕМ**

Сборник научных трудов

Вып. 1 (11), 2016

**Под редакцией чл.-корр. РАН В.И. Борисова**

**Главный редактор А.Г. Остапенко**

Издание публикуется в авторской редакции

Компьютерная верстка Е.Ю. Чапурин

Дизайн обложки С.А. Кравец

Подписано в печать \_\_.12.16. Формат \_\_. Усл. печ. л. \_\_.

Заказ \_\_\_\_ Тираж 170 экз.