

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

Сборник научных трудов

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

Выпуск 2 (12), 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. 2 (12), 2016. – Воронеж: Издательство «Научная книга». 2016. – 105 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

| | |
|---|----|
| СОЦИАЛЬНАЯ СЕТЬ DIGG: СТРУКТУРНО – ФУНКЦИОНАЛЬНЫЙ АНАЛИЗ ПРОЦЕССОВ РАСПРОСТРАНЕНИЯ КОНТЕНТА Кулаков Е.И., Соколова Е.С., Паринов А.В. | 4 |
| СОЦИАЛЬНАЯ СЕТЬ SLASHDOT: СТРУКТУРНО – ФУНКЦИОНАЛЬНЫЕ ОСОБЕННОСТИ РАСПРОСТРАНЕНИЯ КОНТЕНТА Соколова Е.С., Очерет А.В., Паринов А.В. | 27 |
| СТРУКТУРНО – ФУНКЦИОНАЛЬНАЯ СПЕЦИФИКА РАСПРОСТРАНЕНИЯ КОНТЕНТА В СОЦИАЛЬНОЙ СЕТИ RATE&GOODS А.В. Паринов, Д.М. Баранов | 51 |
| СОЦИАЛЬНАЯ СЕТЬ ПО ИНТЕРЕСАМ LAST.FM: СТРУКТУРНО – ФУНКЦИОНАЛЬНЫЕ ОСОБЕННОСТИ АНАЛИЗА ПРОЦЕССОВ РАСПРОСТРАНЕНИЯ КОНТЕНТА Тарелкин М.А., Шварцкопф Е.А., Паринов А.В..... | 71 |
| ШАНСЫ И РИСКИ АТАКУЕМЫХ СЕТЕВЫХ ЭЛЕМЕНТОВ ДЛЯ АНАЛИТИЧЕСКОЙ ОЦЕНКИ ХАРАКТЕРИСТИК ИХ ЖИВУЧЕСТИ Д.Г. Плотников, В.В. Морковина, А.А. Остапенко | 95 |

УДК 004.056

**СОЦИАЛЬНАЯ СЕТЬ DIGG: СТРУКТУРНО –
ФУНКЦИОНАЛЬНЫЙ АНАЛИЗ ПРОЦЕССОВ
РАСПРОСТРАНЕНИЯ КОНТЕНТА**

Кулаков Е.И., Соколова Е.С., Паринов А.В.

Digg.com — новостной социальный сайт, контент которого создается и оценивается пользователями. Его популярность привела к созданию большого числа социальных сайтов с подобной моделью. Данная социальная сеть относится к категории социальных сетей закладок [28].

Чтобы изучить трафик (объем входящей/исходящей из узла сети информации, передаваемой за определённый период времени) в социальной сети Digg, необходимо построить структурно-функциональную модель социальной сети Digg.

Социальная сеть Digg предоставляет различного рода контент (изображения, видеозаписи, текстовая информация, смешанный вид), который обладает своей конкретной ценностью. Эти данные могут нести в себе большое количество информации, вследствие чего пользователи и добавляют их в свои закладки или оценивают [3-6, 12, 13].

Контент, который циркулирует по социальной сети [8-12, 14] включает следующие разновидности (рисунок 1):

1. Текстовый/фото/видео и гибридный контент. Тип контента в зависимости от вида носителя информационного сообщения. В большинстве случаев встречается текстово-фото и текстово-видео контент.

2. Коммерческий/некоммерческий. Коммерческий контент подразумевает разного рода рекламную информацию, в то время как некоммерческий контент представляет собой все новостные сюжеты, информационное интервью и прочее.

Иногда контент коммерческого типа появляется, но выражен неявно, в основном это новостные публикации.

3. Новостной/ профессиональный/ любительский контент. Данному разделению соответствуют репортажи, подготовленные профессиональными журналистами, сотрудниками редакции и читателями ресурса. Любой желающий может публиковаться, вне зависимости от его навыков написания статей.

4. Интерактивный контент подразумевает разнообразные формы активности аудитории такие как комментарии, письма в редакцию, использование социальных плагинов, оценку публикаций. Активность пользователей в данной социальной сети проявляется лишь присвоением Digg'a, в случае если новость понравилась пользователю.

5. Сюжетный и справочный контент. Под сюжетным контентом подразумеваются публикации журналистов и читателей новостного формата, репортажи, художественные произведения. Справочный контент подразумевает информационную справку и т.д.

6. Внутренний и внешний контент. Информация, произведенная непосредственно штатом социальной сети (включая читателей-блогеров), и материалы из внешних источников соответственно.

7. Линейный и гипертекстуальный контент. Линейный контент подразумевает наличие текста в чистом виде, а гипертекстуальный, соответственно, подразумевает наличие ссылок на внутренние или внешние веб-источники.

8. Отдельно хотелось бы выделить магазинный контент. Социальная сеть позволяет приобрести товары разной пригодности.

Именно здесь пользователь может быть атакован посредством XSS. При желании приобрести товар, он может попасть на сайт оплаты злоумышленника и перевести деньги туда, куда захочет сам недоброжелатель.

Поэтому данный контент является одним из опасных видов для невнимательного пользователя.

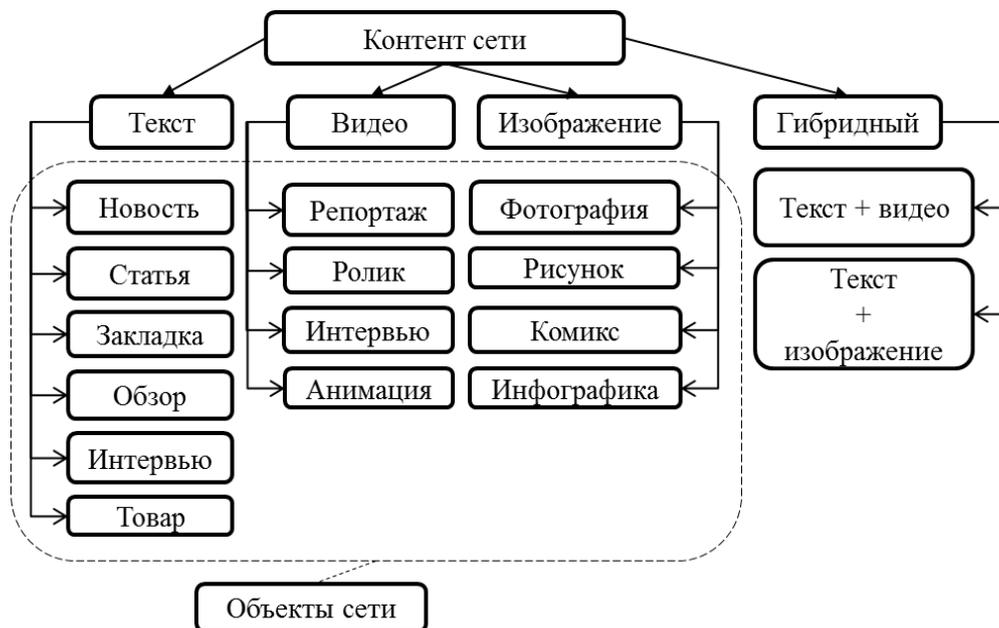


Рис. 1. Контент и объекты в социальной сети Digg

В социальной сети Digg все сетевые ресурсы можно разделить на два основных типа [16, 18]: ресурсы коллективного пользования и ресурсы персонального пользования (рис. 2).

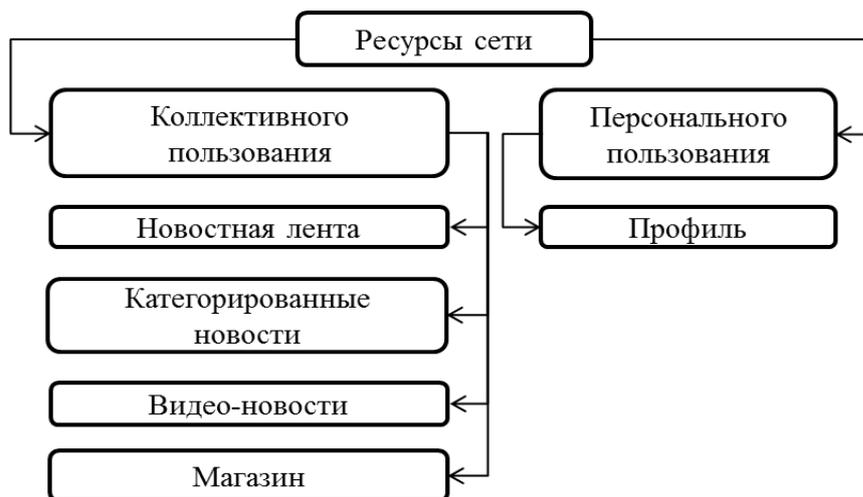


Рис. 2. Ресурсы социальной сети Digg

К ресурсам коллективного пользования можно отнести те ресурсы, которые доступны всем субъектам социальной сети без исключения. Это

может быть новостная лента, видео раздел, новости по категориям или магазин.

К ресурсам персонального пользования можно отнести личный профиль каждого пользователя. В данном случае, доступ к такому ресурсу имеет только непосредственный владелец профиля.

Распишем подробно ресурсы для коллективного и персонального пользования с учетом специфики сети социальных закладок Digg [7, 17]:

Новостная лента - формат данных, используемый для доставки пользователям часто обновляемой информации. Распространители этой информации предоставляют новостную ленту, позволяя пользователям подписаться на неё. Лента состоит из некоторого ограниченного числа статей, а также из служебной информации типа автора ленты или домашней странички. Новости по категориям - подразумевается, что контент определенной области хранится в специально выделенной категории, и в случае чего, пользователю не нужно искать среди всего разнообразия новостей.

Видео-новости - новости, которые содержат в себе видео-контент автоматически попадают в данный сетевой ресурс.

Магазин - предоставляет пользователю возможность приобрести материальный контент, размещенный на странице магазина.

Профиль - хранимая в компьютерной системе совокупность данных о пользователе, необходимая для его опознавания (аутентификации) и предоставления доступа к его личным данным и настройкам. Учётная запись может содержать также дополнительные данные о пользователе.

В социальной сети Digg фигурируют 4 типа ролей пользователей (рисунок 3) [13, 15]:

1. Администратор. Участник сети, создавший группу или публичную страницу и управляющий ролями пользователей.

2. Модератор. Участник сети, который следит за порядком на сетевом пространстве и публикует новости, предложенные авторизованными пользователями.

3. Авторизованные пользователи. Круг лиц, прошедших регистрацию своего аккаунта. К ним относятся авторы и активные пользователи. После регистрации, им можно предлагать новости для публикации, добавлять к себе понравившиеся новости в закладки, а также повышать популярность новости.

4. Неавторизованные пользователи. Оставшаяся масса людей, которым из проявления активности доступно только ознакомление с публикацией.

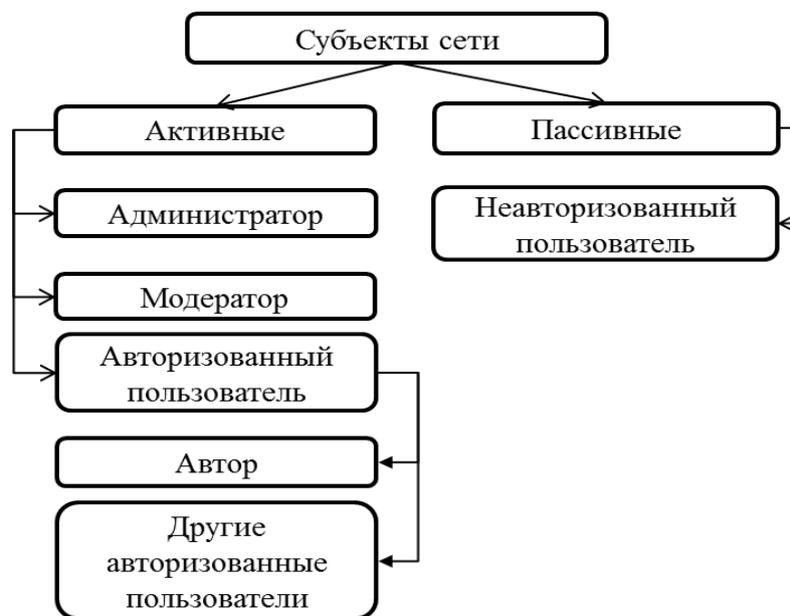


Рис. 3. Субъекты социальной сети Digg

В социальной сети Digg пользователи имеют определенный функционал, позволяющий им взаимодействовать с контентом. Возможности, доступные субъектам данной социальной сети классифицируют по двум признакам [4, 14, 16, 28]: управленческие и персональные (рисунок 4).

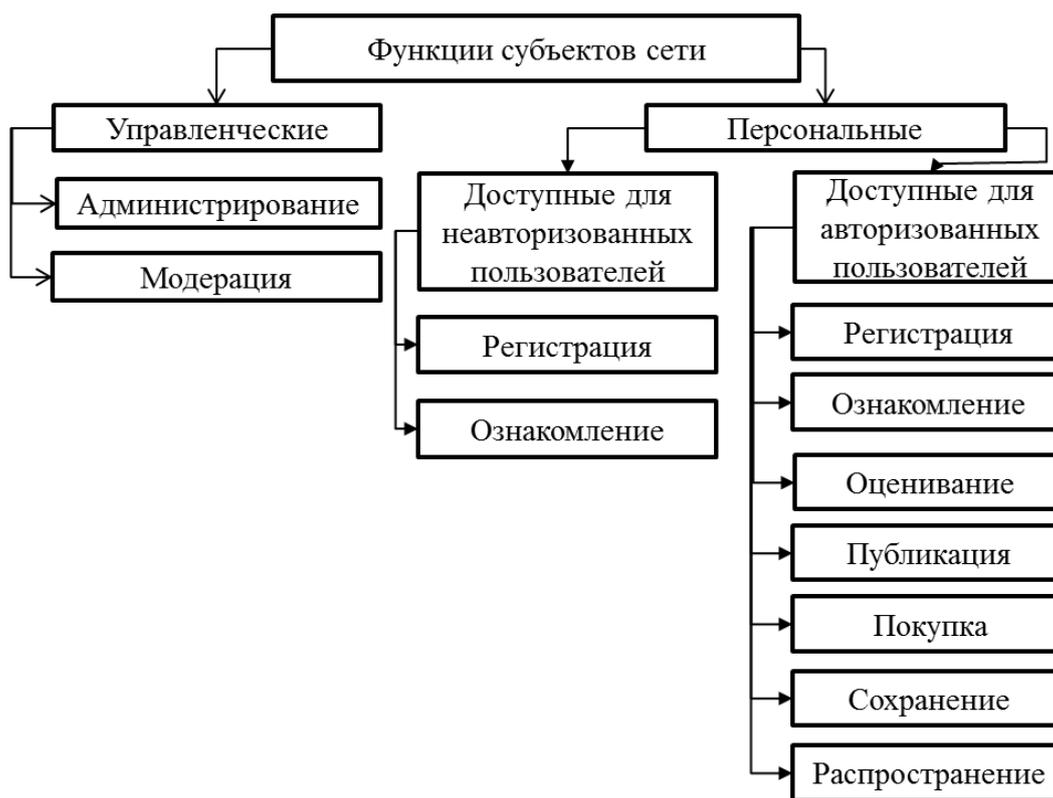


Рис. 4. Функции пользователей социальной сети Diggs

Таким образом, была дана подробная и всесторонняя классификация контента, циркулирующего в социальной сети Diggs. Также, в ходе анализа социальной сети были рассмотрены объекты данной социальной сети и субъекты, которые с ними взаимодействуют. Подробно описаны сетевые ресурсы данной социальной сети. В социальной сети Diggs пользователи имеют определенный функционал, позволяющий им взаимодействовать с контентом.

С учетом полученных классификаций контента, субъектов и их действий, а также сетевых ресурсов данной социальной сети предоставляется возможным построить структурно-функциональную модель социальной сети Diggs с учетом всех ее особенностей (рисунок 5). В данной модели функциональные связи представляют собой сложную структуру взаимодействия контента, сетевых ресурсов и субъектов, функционирующих в заданном сетевом пространстве [10, 12-14].

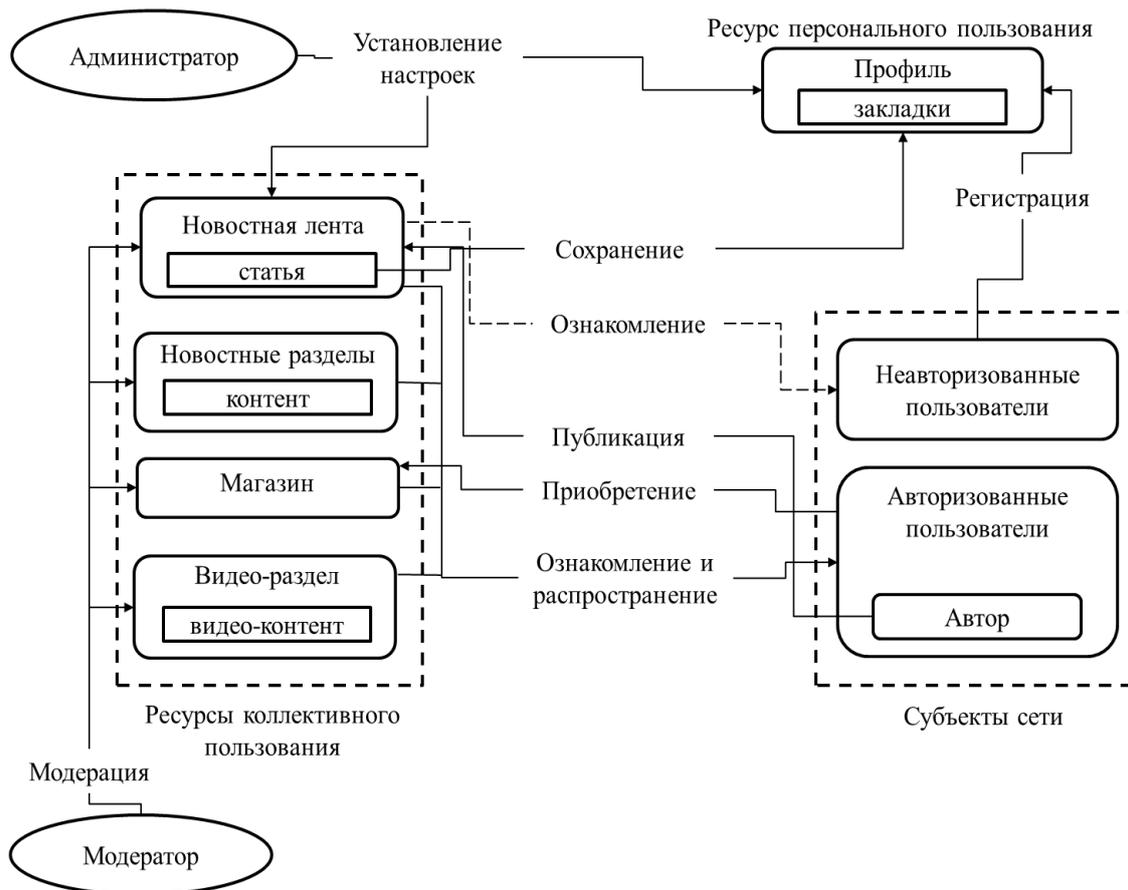


Рис. 5. Структурно-функциональная модель социальной сети Diggs

Рассмотрев структурно – функциональную модель, необходимо перейти к статистическим данным для сети Diggs, которые можно представить в виде трехместного предиката. В нем, на первом месте располагается узел сети, из которого исходит ребро, на втором месте – узел, в который входит это ребро, и на третьем – вес ребра, отражающий силу связи между узлами.

Формату данных трехместного предиката соответствует формула (1):

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

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 .

Фрагмент такого предиката, представленном в текстовом формате, покажем в таблице 1 [9, 10, 12, 16, 21].

Таблица 1 – Фрагмент трехместного предиката социальной сети Digg [20, 29]

| Начальный узел | Конечный узел | Вес |
|----------------|---------------|-----|
| 1 | 2 | 462 |
| 3 | 2 | 837 |
| 4 | 2 | 224 |
| 5 | 2 | 716 |
| 6 | 2 | 965 |
| 7 | 8 | 868 |
| 9 | 8 | 894 |
| 10 | 8 | 491 |
| 11 | 12 | 378 |
| 13 | 12 | 36 |
| 14 | 15 | 37 |

Рассматривать трехместный предикат не является целесообразным, поэтому необходимым элементом для анализа социальных сетей по интересам является нахождение взвешенных метрик.

Вершины трехместного предиката означают субъектов анализируемой социальной сети: администраторы, модераторы, авторы, авторизованные или неавторизованные пользователи. Взаимодействуя между собой, субъекты связываются между собой силой связи, называемой весом. Эта сила определяется функциональной зависимостью между субъектами. Вес дуги означает количество сохраненных закладок за единицу времени (единица времени - год) [12, 14].

Трехместный предикат позволяет построить звездную матрицу, элементы строки которой соответствуют дугам, входящим в данную вершину, а элементы столбца – дугам, исходящим из вершины. Данная матрица будет квадратная, размерности $N \times N$, где N - количество вершин в сети, а диагональные элементы будут равны нулю. Данную звездную матрицу покажем в таблице 2.

Таблица 2 – Звездная матрица взвешенности для сети Digg

| | | | | | | | | |
|-------|-----|-----|-----|-----|-----|-----|-----|-------|
| | 1 | 2 | ... | 11 | ... | 21 | ... | 30398 |
| 1 | 0 | 462 | ... | 0 | ... | 0 | ... | 0 |
| 2 | 0 | 0 | ... | 0 | ... | 0 | ... | 0 |
| ... | ... | ... | ... | ... | ... | ... | ... | ... |
| 11 | 0 | 0 | ... | 0 | ... | 0 | ... | 0 |
| ... | ... | ... | ... | ... | ... | ... | ... | ... |
| 21 | 0 | 0 | ... | 0 | ... | 0 | ... | 0 |
| ... | ... | ... | ... | ... | ... | ... | ... | ... |
| 30398 | 0 | 0 | ... | 0 | ... | 0 | ... | 0 |

Для дальнейшего моделирования процессов диффузии контента необходимо для каждой вершины сети знать количество исходящих из нее дуг, которое назовем ее степенью исхода. Рассчитав степень исхода для каждой вершины, построим диагональную матрицу степени вершин сети (таблица 3). В ней по диагонали располагаются значения степеней для каждой вершины, рассчитанные исходя из трехместного предиката.

Таблица 3 – Диагональная матрица степени вершин для сети Digg

| | | | | | | | | | | |
|-------|-----|-----|-----|-----|-----|-----|-----|-----|-----|-------|
| | 1 | 2 | 3 | ... | 11 | 12 | ... | 21 | ... | 30398 |
| 1 | 1 | 0 | 0 | ... | 0 | 0 | ... | 0 | ... | 0 |
| 2 | 0 | 5 | 0 | ... | 0 | 0 | ... | 0 | ... | 0 |
| 3 | 0 | 0 | 2 | ... | 0 | 0 | ... | 0 | ... | 0 |
| ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... |
| 11 | 0 | 0 | 0 | ... | 4 | 0 | ... | 0 | ... | 0 |
| ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... |
| 21 | 0 | 0 | 0 | ... | 0 | 0 | ... | 5 | ... | 0 |
| ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... |
| 30398 | 0 | 0 | 0 | ... | 0 | 0 | ... | 0 | ... | 2 |

После, построим квадратную матрицу взвешенной центральности элементов сети. Для этого определяем удельный вес ее вершин и дуг с помощью нормировки их весов по суммарному трафику сети по формуле 2. Данная формула (2) характеризует степень взвешенной (по трафику) центральности дуги.

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

где $\delta(a_{ij})$ – вес трафика в дуге a_{ij} .

Заметим, что исходя из суммарного трафика сети, не происходит разделение на входящие и исходящие дуги. Поэтому, для определения взвешенной центральности вершины x_s используем сумму, которую сразу нормируем по суммарному трафику сети.

В результате получаем нормированную величину (формула 3), которая характеризует удельный вес трафика, проходящего через вершину x_s , по отношению ко всему трафику сети:

$$\delta(\bar{x}_s) = \left[\sum_i \delta(a_{si}) + \sum_j \delta(a_{js}) \right] / \sum_{\substack{i,j \\ i \neq j}} \delta(a_{ij}), \quad (3)$$

где $\delta(a_{si})$ – вес трафика в исходящей из узла a_{si} дуге;

$\delta(a_{js})$ – вес трафика во входящей в узел a_{js} дуге.

В итоге, используя формулу 3, получим матрицу взвешенной центральности элементов исходной сети следующего вида: в ней по диагонали располагаются значения удельного веса для каждой вершины, остальные вершины равны 0. Изобразим полученную матрицу в таблице 4.

Таблица 4 – Матрица взвешенной центральности для сети Digg

| | | | | | | | | |
|-------|---------|---------|-----|----------|-----|---------|-----|----------|
| | 1 | 2 | ... | 11 | ... | 21 | ... | 30398 |
| 1 | 0,00042 | 0,00012 | ... | 0 | ... | 0 | ... | 0 |
| 2 | 0 | 0,00027 | ... | 0 | ... | 0 | ... | 0 |
| ... | ... | ... | ... | ... | ... | ... | ... | ... |
| 11 | 0 | 0 | ... | 0,000851 | ... | 0 | ... | 0 |
| ... | ... | ... | ... | ... | ... | ... | ... | ... |
| 21 | 0 | 0 | ... | 0 | ... | 0,00055 | ... | 0 |
| ... | ... | ... | ... | ... | ... | ... | ... | ... |
| 30398 | 0 | 0 | ... | 0 | ... | 0 | ... | 0,000228 |

В конечном итоге наших преобразований, построим диагональную матрицу удельного баланса трафика в вершинах социальной сети Digg.

Здесь, в отличие от взвешенной центральности, учитывается не сумма, а разность (формула 4) исходящего и входящего для вершины трафика:

$$\delta(\bar{B}_s) = \left[\sum_i \delta(a_{si}) - \sum_j \delta(a_{js}) \right] / \sum_{\substack{i,j \\ i \neq j}} \delta(a_{ij}), \quad (4)$$

где $\delta(a_{si})$ – вес трафика в исходящей из узла a_{si} дуге;

$\delta(a_{js})$ – вес трафика во входящей в узел a_{js} дуге.

В результате вычислений построим диагональную матрицу удельного баланса трафика в вершинах сети (таблица 5). Ее диагональные элементы характеризуют роль вершины (субъекта) сети и, чем больше модуль \bar{B}_s , тем в большей степени проявляется его ролевая функция:

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

Имея диагональную матрицу удельного баланса трафика, мы нашли те субъекты, которые публикуют контент, и те, которые его потребляют.

Таблица 5 - Диагональная матрица удельного баланса трафика

| | | | | | | | | |
|-------|-----------|-----------|-----|------------|-----|-----------|-----|----------|
| | 1 | 2 | ... | 11 | ... | 22 | ... | 30398 |
| 1 | -0,000010 | 0 | ... | 0 | ... | 0 | ... | 0 |
| 2 | 0 | -0,000034 | ... | 0 | ... | 0 | ... | 0 |
| ... | ... | ... | ... | ... | ... | ... | ... | ... |
| 11 | 0 | 0 | ... | -0,0000545 | ... | 0 | ... | 0 |
| ... | ... | ... | ... | ... | ... | ... | ... | ... |
| 22 | 0 | 0 | ... | 0 | ... | 0,0000236 | ... | 0 |
| ... | ... | ... | ... | ... | ... | ... | ... | ... |
| 30398 | 0 | 0 | ... | 0 | ... | 0 | ... | 0,000030 |

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 [16-20].

Поэтому необходимо найти репрезентативную выборку, которая имеет значительно меньшую размерность, но с заданной точностью сохраняет свойства исходной сети.

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

Поэтому будем использовать диагональ квадратной матрицы взвешенной центральности элементов сети, где определена степень взвешенной центральности каждой вершины. Данный подход является объективным, потому что учитывает трафик – основу функционирования любой сети.

Изначально выполним вычленение из квадратной матрицы взвешенной центральности элементов сети ее диагонали (таблица 6), в которой будут сосредоточены только степени взвешенной центральности ее вершин.

Таблица 6 – Степени взвешенной центральности вершин

| Номер узла | Степень |
|------------|----------|
| 1 | 0,000429 |
| 2 | 0,000274 |
| 3 | 0,000168 |
| ... | ... |
| 21 | 0,000559 |
| 22 | 0,000820 |
| 23 | 0,000152 |
| ... | ... |
| 30398 | 0,000228 |

Далее пересортируем степени взвешенной центральности вершин в порядке убывания (таблица 7), тем самым найдем самые центральные узлы

в нашей сети. Исходя из таблицы 7, можно найти те субъекты сети, через которые проходит большая часть трафика.

Таблица 7 – Ранжированные степени взвешенной центральности вершин

| Номер узла | Степень |
|------------|----------|
| 8054 | 0,01422 |
| 342 | 0,01370 |
| 6482 | 0,01162 |
| ... | ... |
| 20573 | 0,00985 |
| 16752 | 0,00957 |
| ... | ... |
| 441 | 0,000052 |

Просуммируем значения (формула 5) из таблицы 7 (начиная с первого) для получения необходимого результата (точности модели). т.е. до значения 0.95, так как допустима 5% потеря трафика. Этим критерием (доверительным интервалом) и ограничится репрезентативная (с точки зрения трафика) выборка [18]:

$$\delta(\bar{x}_s) = 0,01422 + 0,01370 + \dots + 0,00957 + \dots + 0,000052 = 0,951. \quad (5)$$

При данном суммировании получили выборку из 379 элементов.

Далее удалим 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 (таблица 8) будет полностью отражать связность компонентов выборки. Как описывалось ранее, ее диагональные значения будут равны нулю.

Таблица 8 – Проранжированная усеченная квадратная матрица

| | | | | | | | | | |
|-------|---------|---------|---------|-----|---------|---------|---------|-----|-----|
| | 8054 | 342 | 6482 | ... | 20573 | 16752 | 947 | ... | 441 |
| 8054 | 0 | 0 | 0,01370 | ... | 0 | 0 | 0 | ... | 0 |
| 342 | 0,01422 | 0 | 0 | ... | 0 | 0 | 0 | ... | 0 |
| 6482 | 0 | 0,01162 | 0 | ... | 0 | 0 | 0 | ... | 0 |
| ... | ... | ... | ... | ... | ... | ... | 0 | ... | ... |
| 20573 | 0 | 0 | 0 | ... | 0 | 0,00957 | 0 | ... | 0 |
| 16752 | 0 | 0 | 0 | ... | 0,00985 | 0 | 0 | ... | 0 |
| 947 | 0 | 0 | 0 | | 0 | 0,00842 | 0 | | 0 |
| ... | ... | ... | ... | ... | ... | ... | ... | ... | ... |
| 4184 | 0 | 0 | 0 | | 0 | 0 | 0,00578 | ... | 0 |
| 441 | 0 | 0 | 0 | ... | 0 | 0 | 0 | ... | 0 |

В конечном итоге сформируем итоговую матрицу (таблица 9) в виде, удобном для последующего моделирования информационной диффузии.

Представляется, что она должна выглядеть следующим образом:

- недиагональные элементы равны (± 1), так как интересует лишь факт смежности вершин истоков (+1) и стоков (-1);

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

Таблица 9 - Модифицированная матрица смежности

| | | | | | | | | | | |
|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|
| | 1 | 2 | ... | 11 | 12 | ... | 21 | 22 | ... | 379 |
| 1 | 4 | 1 | ... | 0 | 0 | ... | 0 | 0 | ... | 0 |
| 2 | -1 | 8 | ... | 0 | 0 | ... | 0 | 0 | ... | 0 |
| ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... |
| 11 | 0 | 0 | ... | 2 | 1 | ... | 0 | 0 | ... | 0 |
| ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... |
| 21 | 0 | 0 | ... | 0 | 0 | ... | 15 | 0 | ... | 0 |
| ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... |
| 379 | 0 | 0 | ... | 0 | 0 | ... | 0 | 0 | ... | 7 |

Отсюда определяется ожидаемое количество зараженных контентом элементов. При этом надо понимать, что это весьма упрощенное представление и, при необходимости учета заметности контента на фоне другого трафика дуги, нужно учитывать ее трафик.

Выполнив вычисления для получения репрезентативной выборки из сети и имея размер выборки, мы получим самые центральные вершины, усеченную матрицу взвешенности, усеченную матрицу взвешенной центральности, с помощью которых можем анализировать усеченную топологию нашей исследуемой сети [16].

Для того, чтобы с уверенностью пользоваться полученной выборкой, необходимо доказать ее репрезентативность. Одним из способов доказать подобие основной сети и выборки, полученной в результате применения алгоритма (на основе доказательства о принадлежности полученной выборки сети к степенному закону распределения), является критерий Пирсона [19].

Посчитаем критерий Пирсона для выборки сети Digg по формуле 6:

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

где n – объем выборки;

p_i - вероятность попадания случайной величины X .

Вычислим вероятности попадания случайной величины X в частичные интервалы $[x_i, x_{i+1}]$ для Digg. Формула (7) расчета вероятностей для степенного распределения:

$$p_i = \int_k^{k+1} ck^{-\alpha} dk = c \left(\frac{k^{1-\alpha}}{\alpha-1} + \frac{(k+1)^{1-\alpha}}{1-\alpha} \right). \quad (7)$$

Возьмем за k количество вершин для выбранной степени вершины.

Пусть $c = 1$, $\alpha = 2.18$, тогда:

$$p_1 = 0.000039; p_2 = 0.000056;$$

$$p_3 = 0.000100; p_4 = 0.000140;$$

$$p_5 = 0.000185; p_6 = 0.000268;$$

$$p_7 = 0.000474; p_8 = 0.000728;$$

$$\begin{aligned}
p_9 &= 0.000937; p_{10} = 0.001128; \\
p_{11} &= 0.001383; p_{12} = 0.001542; \\
p_{13} &= 0.001730; p_{14} = 0.001952; \\
p_{15} &= 0.002220; p_{16} = 0.002943; \\
p_{17} &= 0.004069; p_{18} = 0.004882; \\
p_{19} &= 0.004882; p_{20} = 0.005956; \\
p_{21} &= 0.964384.
\end{aligned}$$

Берем $n = 21$, тогда np_i будут равны:

$$\begin{aligned}
np_1 &= 0,000816; np_2 = 0,001168; \\
np_3 &= 0,002091; np_4 = 0,002950; \\
np_5 &= 0,003895; np_6 = 0,005628; \\
np_7 &= 0,009948; np_8 = 0,015298; \\
np_9 &= 0,019681; np_{10} = 0,023698; \\
np_{11} &= 0,029033; np_{12} = 0,032380; \\
np_{13} &= 0,036320; np_{14} = 0,041002; \\
np_{15} &= 0,046619; np_{16} = 0,061806; \\
np_{17} &= 0,085459; np_{18} = 0,102529; \\
np_{19} &= 0,102529; np_{20} = 0,125074; \\
np_{21} &= 20,25207.
\end{aligned}$$

Используя формулу 6, получаем $\chi^2=22.05426$.

Выберем уровень значимости $\alpha=0,05$, количество степеней свободы равно $(n - 1) = 20$, тогда $p = 1 - \alpha = 0.95$. Для данных параметров получаем $\chi^2(20) = 22.05426$.

Табличное значение $\chi^2(20)$ равно 31.4. Получается, что табличное значение χ^2 больше вычисленного. Следовательно, приведенное распределение является степенным, что и требовалось доказать [19].

Далее, рассчитаем среднеквадратичные отклонения (СКО) для наших выборок и найдем относительное значение, полученных отклонений.

Расчёт СКО осуществим с применением формулы 8 и 9 [19]:

$$\bar{X} = \frac{\sum_{i=1}^n X_i}{n}; \quad (8)$$

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

где X_i – элементы совокупности;

\bar{X} – среднее арифметическое выборки.

По формуле 9 рассчитали СКО генеральной совокупности $\sigma_1 = 0.293$ и СКО выборки $\sigma_2 = 0.2785$.

Отклонение между эталоном (генеральной совокупностью) и полученной выборкой получилось равным 4.94%. Это доказывает, что генеральная совокупность и выборка подобны в пределах допустимого значения 5% [20].

Доказательство сохранения всех свойств было произведено с помощью критерия Пирсона, показавшего, что полученная выборка также подчиняется степенному закону распределения, следовательно, исходная и вновь полученная сети являются подобными.

Далее представим гистограмму распределения вершин для первоначальной и усеченной сетей (по горизонтальной оси – степень вершины, по вертикальной – доля вершин степени k) на рисунке 6.

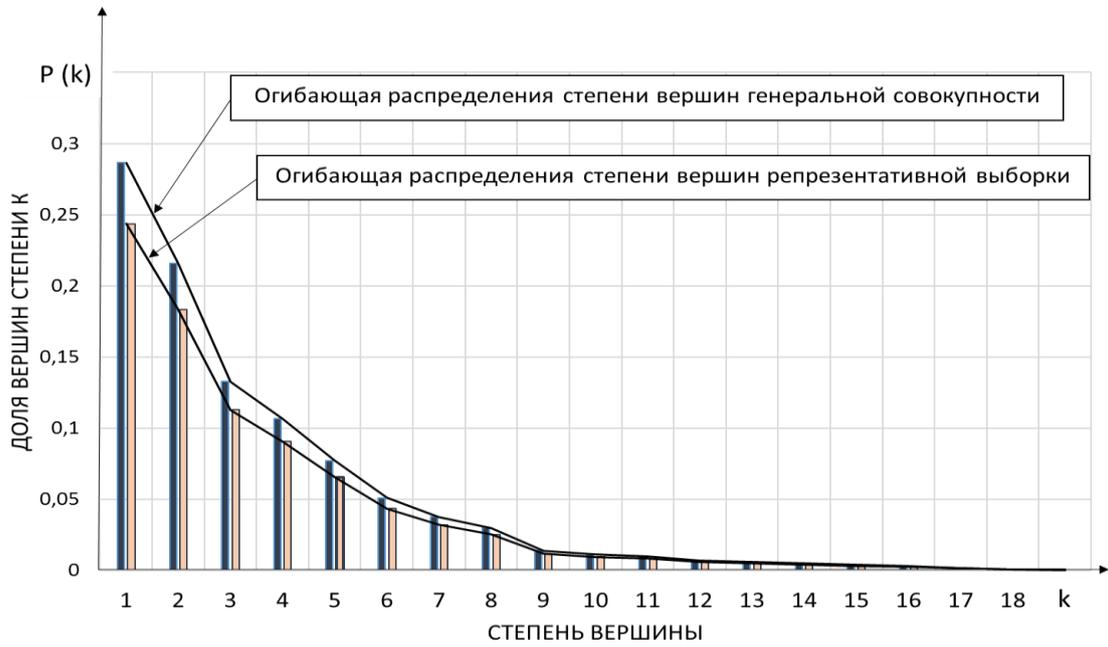


Рис. 6. Гистограмма распределения вершин Digg

После применения алгоритма свертки, покажем изображение усеченной топологии сети на рисунке 7.

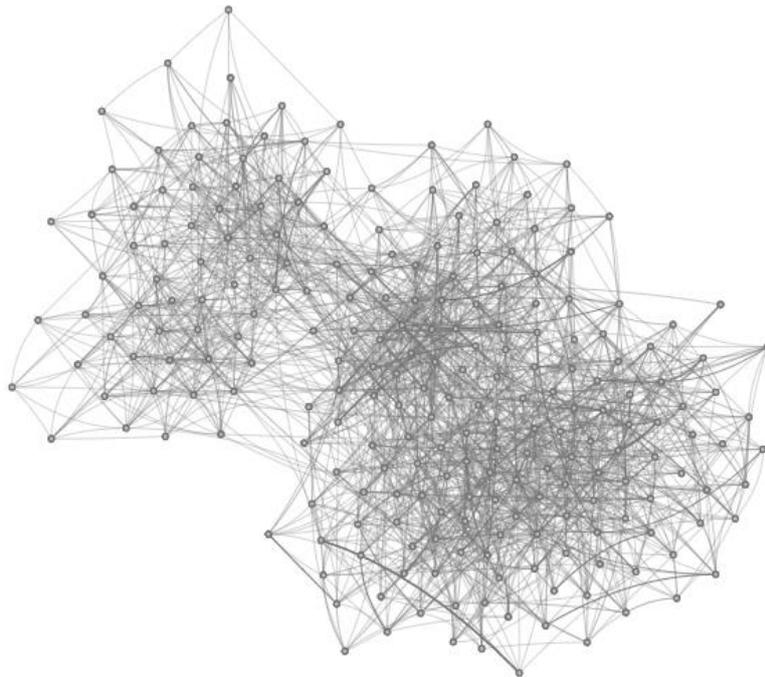


Рис. 7. Усеченная топология социальной сети Digg

Здесь, в отличии от одного большого кластера основной топологии, видим, что усеченная топология представлена пятью, меньшего размера, кластерами. Они означают активность (сохранение, распространение, оценивание и т.п.) пользователей в пяти категориях новостей, имеющихся в социальной сети Digg. Также проанализировав граф сжатой сети Digg, можно отчетливо различить в пределах каждого кластера данной сети вероятные хабы-генераторы вредоносного трафика, которые обладают наибольшим количеством каналов для передачи вредоносного контента в сеть, что является опасным с точки зрения информационной безопасности, так как за крайне короткий срок в сеть может быть вброшено огромное количество информации, что повлечет за собой крайне динамично развивающуюся эпидемию заражения [22-27].

Таким образом, в ходе анализа были выполнены алгоритмы преобразования исходных данных сети и нахождения репрезентативной выборки, получена визуальная модель исследуемой сети, а также вычислены соответствующие матрицы, позволяющие провести дальнейший анализ с целью получения полной картины распространения контента в социальной сети. Доказана репрезентативность выборки генеральной совокупности с помощью критерия согласия Пирсона, найдено среднеквадратичное отклонение выборки в 5% от генеральной совокупности и графическим методом показано подобие выборки.

Литература:

1 Alan E. Mislove. Online Social Networks: Measurement, Analysis, and Applications to Distributed Information Systems. Houston, Texas: RICE University, 2009. – P. 206.

2 Alba R.A., graph-theoretic definition of a sociometric clique / Richard D. Alba / Journal of Mathematical Sociology. – 2013. – P. 113–126.

3 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.

4 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.

5 Anthonisse J. M., The rush in a directed graph / J.M. Anthonisse // Technical Report BN 9/71. – 1971 p. Ahn Y. Analysis of topological characteristics of huge online social networking services / Y. Ahn, S. Han, H. Knak, S. Moon, H. Jeong // 16th International Conference on the World Wide Web. – 2014. – P. 835-844

6 Antsupov, A. Ya. Conflictology: the textbook for higher education institutions / A.Ya. Antsupov, A. I. Shilov. – 3rd prod., revised and additional – SPb.: St. Petersburg, 2007. – 591 p.

7 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.

8 Discrete 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.

9 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.

10 G. Szabó and B. A. Huberman, “Predicting the popularity of online

content” Commun. ACM, vol. 53, no. 8, pp. 80–88, Aug. 2015.

11 Haythornthwaite C. 2015. Social networks and internet connectivity effects. *Info, Communication & Society*, 8(2), – P. 125–147.

12 K. Lerman, “Social networks and social information filtering on Digg,” in Proc. 1st Int. Conf. Weblogs and Social Media (ICWSM-07), 2014.

13 Kalashnikov, A.O. Attacks to information and technological infrastructure of crucial objects: assessment and regulation of risks: Monograph / A.O. Kalashnikov, E. V. Yermilov, O. N. Choporov, K. A. Razinkin, N. I. Barannikov; under the editorship of the Member correspondent of RAS D. A. Novikov. - Voronezh: Scientific Book publishing house. 2013. - 160 p.

14 M. Cha, H. Kwak, P. Rodriguez, Y. Y. Ahn, and S. Moon, “Analyzing the video popularity characteristics of large-scale user generated content in social network Digg,” *IEEE/ACM Trans. Netw.*, vol. 17, no. 5, pp. 1357–1370, 2015.

15 Newman, M. E. Finding and evaluating community structure in networks / M. E. J. Newman, M. Girvan // *Phys. Rev. E* 69. – 2014 – P. 53-58.

16 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.

17 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 (Spl.End). – P. 251-255.

18 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.

19 Гмурман В.Е., Теория вероятностей и математическая статистика. Учебное пособие. Высшее образование. – Москва, 2006 – С. 243.

20 Губанов Д.А., Новиков Д.А., Чхартишвили А.Г. Социальные сети: модели информационного влияния, управления и противоборства / Под ред. чл.-корр. РАН Д.А. Новикова. –М.: Издательство физико-математической литературы, 2010. – 116с.

21 Остапенко Г.А. Алгоритмизация оценки живучести сетевых информационных структур [Текст] / Г.А. Остапенко, Я.С. Мишина, В.И. Белоножкин, И.В. Шевченко // Информация и безопасность. – 2014. – Т. 17. – № 2. – С. 304-307.

22 Паринов А.В. Риск-оценка смертности инновационных проектов: научно-методические основы [Текст]/ А.В. Паринов, Л.В. Паринова, В.Г. Юрасов // Информация и безопасность. – 2013. – Т. 16. – № 3. – С. 423-426.

23 Паринов А.В. Управление рисками инновационных проектов [Текст]/ А.В. Паринов, Л.В. Паринова, А.В. Заряев // Информация и безопасность. – 2014. – Т. 17. – № 4. – С. 602-605.

24 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.

25 Поликарпов Э.Д. Методическое обеспечение учебного процесса подготовки специалистов в области обеспечения информационной безопасности телекоммуникационных систем: виртуальные лабораторные работы по исследованию средств приема и передачи цифровых сигналов [Текст]/ Э.Д. Поликарпов, В.А. Манюхин, В.Н. Деревянко // 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. 7. – № 1. – С. 67-75.

26 Радько Н.М. Топологическое многообразие эпидемических моделей неоднородных информационно-телекоммуникационных сетей в контексте их эпистойкости [Текст]/ Н.М. Радько, В.В. Исламгулова, Д.Н. Рахманин // 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. 54-65.

27 Романов С.К. Переходные процессы в синтезаторе частот с одновременно коммутируемыми трактами приведения частоты и каналами управления [Текст]/ С.К. Романов, Н.М. Тихомиров, А.В. Гречишкин, Д.Н. Рахманин, В.Н. Тихомиров // Вестник Московского государственного технического университета им. Н.Э. Баумана. Серия: Приборостроение. – 2014. – Т. 8. – № 2. – С. 76-93.

28 Социальная сеть Digg. - Электрон. Дан. – Режим доступа: <http://digg.com> (дата обращения: 06.08.2016)

29 Средство визуализации данных. - Электрон. Дан. – Режим доступа: <https://gerphi.org/> (дата обращения: 15.09.2016)

30 Статистика социальной сети Digg. - Электрон. Дан. – Режим доступа: <http://konect.uni-koblenz.de/networks/digg> (дата обращения: 07.09.2016).

Воронежский научно-образовательный центр
управления информационными рисками

**СОЦИАЛЬНАЯ СЕТЬ SLASHDOT:
СТРУКТУРНО – ФУНКЦИОНАЛЬНЫЕ ОСОБЕННОСТИ
РАСПРОСТРАНЕНИЯ КОНТЕНТА**

Соколова Е.С., Очерет А.В., Паринов А.В.

Современный мир характеризуется большой информационной конфронтацией. Не исключением стали и социальные сети, которые с каждым днем все больше охватывают сферы нашего существования. В связи с этим количество реализуемых угроз, которые возможны в социальных сетях, такие как спам, информационно-психологическое воздействие, и в целом распространение вредоносного контента постоянно растет, что говорит о широких возможностях для злоумышленников по осуществлению различных деструктивных воздействий [1-3].

Социальная сеть Slashdot представляет собой англоязычный новостной сайт, специализирующийся на технических и интересных технической аудитории темах [4, 5]. Он предлагает пользователям отправлять новости с соответствующими ссылками на источник о компьютерном оборудовании, девайсах, играх, облаке, мобильных устройствах, хранении информации, безопасности, управлении данными, книжных обзорах и т.д. Все материалы публикуются после проверки редакцией сайта (модераторами). В данной социальной сети предложенные материалы для публикаций проходят проверку пятью модераторами и после этого публикуются в соответствующем новостном разделе [4].

Для подробного и всестороннего анализа структурно-функциональной особенности социальной сети Slashdot необходимо, в первую очередь, провести классификацию ресурсов (рис. 1), контента, циркулирующего в заданном сетевом пространстве (рис. 2), объектов и

субъектов социальной сети (рис. 2, 3), а также установить функциональные взаимосвязи между всеми элементами сети (рис. 4, 5).

Сетевые ресурсы можно представить в виде следующей схемы (рис. 1):



Рис. 1. Классификация сетевых ресурсов социальной сети Slashdot

Профиль пользователя – является ресурсом личного пользования. В нем пользователи могут сохранять закладки, производить настройку новостной ленты и выкладывать личную информацию (контакты).

Новостная лента – формат данных, используемый для доставки пользователям часто обновляемой информации. Распространители этой информации предоставляют новостную ленту, позволяя пользователям подписаться на неё.

Новостные разделы – подразумевается, что весь контент области хранится в специально выделенном разделе на сайте, что упрощает навигацию по сайту и поиск нужной пользователю статьи.

Главная страница сайта Slashdot – страница, которую видит любой (авторизованный или неавторизованный) пользователь, который зашел на

данный сайт. На этой странице отображаются недавно опубликованные новостные статьи, реклама и навигация по социальной сети [4, 5].

Классификацию контента целесообразно рассматривать с точки зрения его представления в социальной сети Slashdot.

Для данной социальной сети представление контента актуально в четырех формах: текстовом, видео, графическом и гибридном.

Под гибридным контентом следует понимать контент, полученный вследствие сочетания (двух или более) в себе различающихся форм представления информации [4, 5].

Весь контент, циркулирующий в данной социальной сети, можно разделить на различные объекты, создаваемые и публикуемые пользователями. Все объекты актуальные для данной социальной сети были разделены по соответствующему контенту (рис. 2).

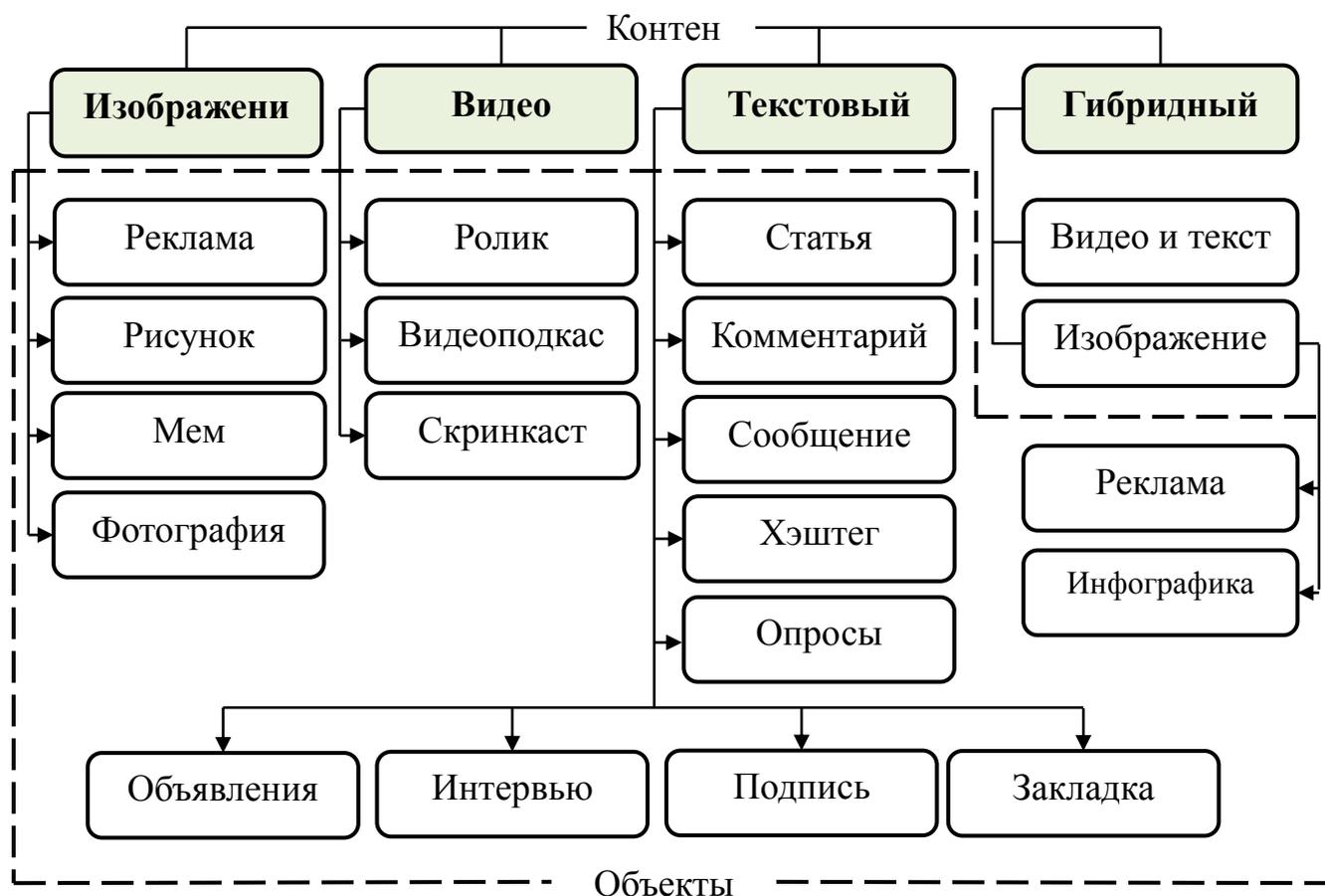


Рис. 2. Классификация контента и объектов в социальной сети Slashdot

При анализе аудитории (рис. 3) данной социальной сети целесообразно выделить три основных типа субъектов задействованных в публикации, просмотре, комментировании и обсуждении тех или иных статей в соответствующих разделах [1-3].

К первому типу можно отнести модераторов или администраторов сетевых ресурсов. Данные субъекты имеют одинаковые права, следят за порядком в социальной сети и занимаются проверкой публикуемых новостных статей или любого другого контента, который поддерживается в данной социальной сети [1, 2, 6].

Второй тип субъектов представлен авторизованными пользователями. В данный перечень входят пользователи социальной сети, прошедшие процедуру регистрации аккаунта. После данной процедуры субъекты социальной сети получают расширенные возможности. В свою очередь, весь блок авторизованных пользователей можно разделить на две части: активных и пассивных пользователей.

– Активные пользователи. Данная часть пользователей участвует в дискуссиях (комментировании) заинтересовавшей их статье, обсуждениях интересующих их вопросов в личных блогах, предложением собственных статей, оценке других пользователей на предмет адекватности высказанного убеждения или мнения в дискуссии. Также данная часть авторизованных пользователей может поделиться предложенной статьей в другой социальной сети (Facebook, Google+ и другие), тем самым привлекая внимание как к самой статье, других участников сетевого пространства [1, 2, 6].

– Пассивные пользователи, которые не принимают активного участия в дискуссиях и оценке других пользователей. Их не интересует публикация собственных статей и ведение блогов. Они знакомятся с новыми статьями на главной странице социальной сети или с помощью новостной ленты. [1, 2, 6].

К последнему типу субъектов представленных в данной социальной

сети можно отнести неавторизованных пользователей. Это оставшаяся масса людей, которые имеют ограниченные возможности в данной социальной сети. Им доступен только просмотр статей на главной странице сайта и возможность поделиться статьей в другой социальной сети [1, 2, 6].

Представим данную классификацию на следующем рисунке (рис. 3).

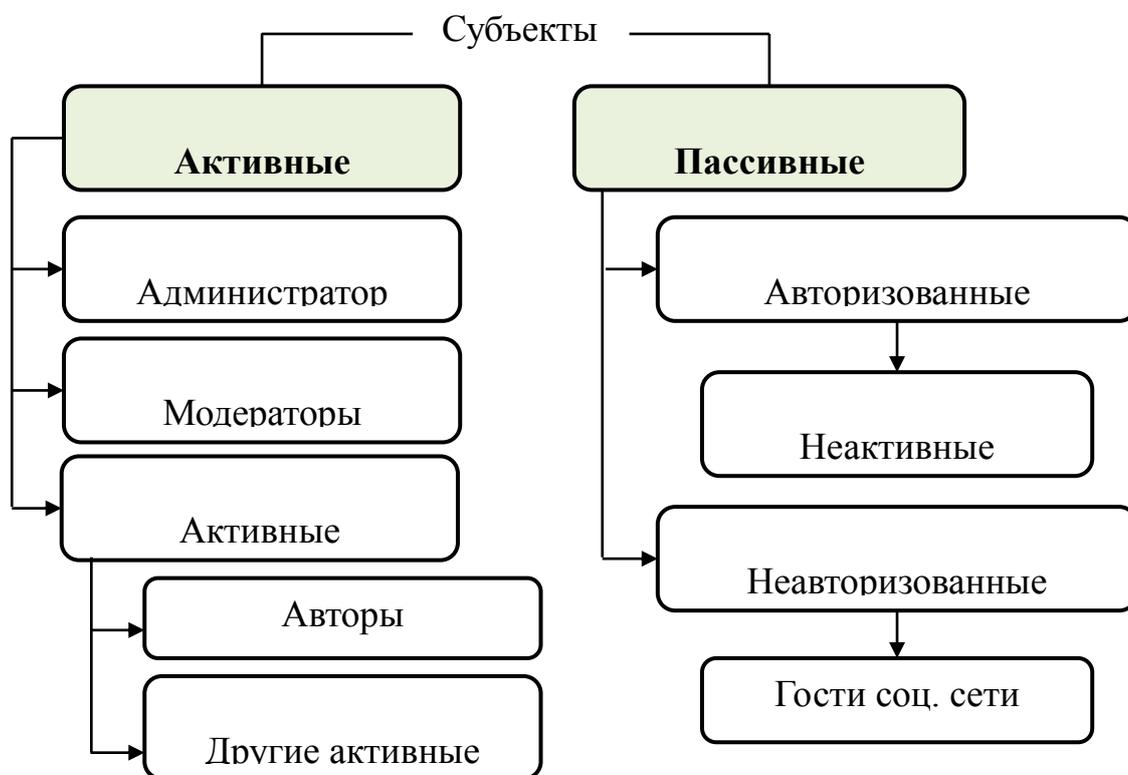


Рис. 3 – Классификация субъектов социальной сети Slashdot

В социальной сети Slashdot пользователи имеют определенный набор действий (функций), позволяющий им взаимодействовать с контентом. Все возможности доступные субъектам данной социальной сети целесообразно классифицировать по функциональной возможности (рис. 4).

В данной классификации действия доступные пользователям были разделены на три основные категории: размещение, ознакомление и реагирование.

В категории «размещение», пользователям доступно только одно действие – публикация.

В категории «ознакомление», пользователи способны прочитать статью или посмотреть видео, рекламу или фотографию.

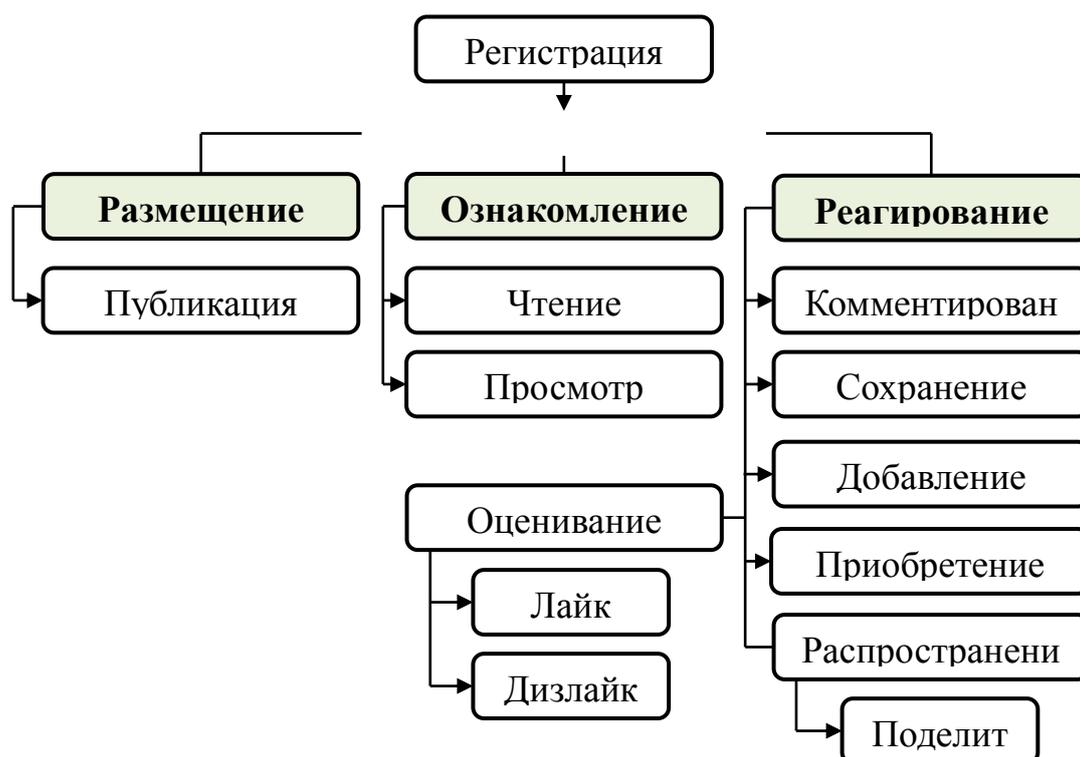


Рис. 4. Классификация действий субъектов в социальной сети Slashdot с позиции их функционала

Весь функционал пользователя социальной сети Slashdot сосредоточен в категории «реагирование». В данную категорию попали действия, которые следуют после ознакомления пользователей с предложенным контентом.

Проведенная классификация способна дать представление о полном списке возможностей всех субъектов социальной сети Slashdot, а также построить структурно-функциональную модель социальной сети Slashdot.

С учетом полученных классификаций контента, субъектов и их действий, а также сетевых ресурсов данной социальной сети предоставляется возможным построить структурно-функциональную модель социальной сети Slashdot с учетом всех ее особенностей (рис. 5).

В данной модели функциональные связи представляют собой сложную структуру взаимодействия контента, сетевых ресурсов и субъектов, функционирующих в заданном сетевом пространстве.

Вышеизложенные классификации (рис. 1-4) и структурно-функциональная модель (рис. 5), в дальнейшем позволят проанализировать действия и поведение субъектов социальной сети в различных ситуациях, а также разработать математические микро-модели индивидуального информационного воздействия на пользователя на основе риск-анализа возникновения социальных ущербов, включая расчет вероятности прочтения, восприятия, одобрения и распространения деструктивного контента, побуждающему к антисоциальным действиям, таким как террор, протесты против конституционного строя и т.п.

Полученные микро-модели позволят в полной мере описать и показать распространение вредоносного контента в сети в зависимости от его популярности, ценности и других немаловажных факторах, которые следует учесть при построении модели эпидемии.

Затем, на основе полученных моделей эпидемий следует выработать рекомендации и разработать методики противодействия атакующему деструктивному контенту в информационном пространстве социальной сети Slashdot.

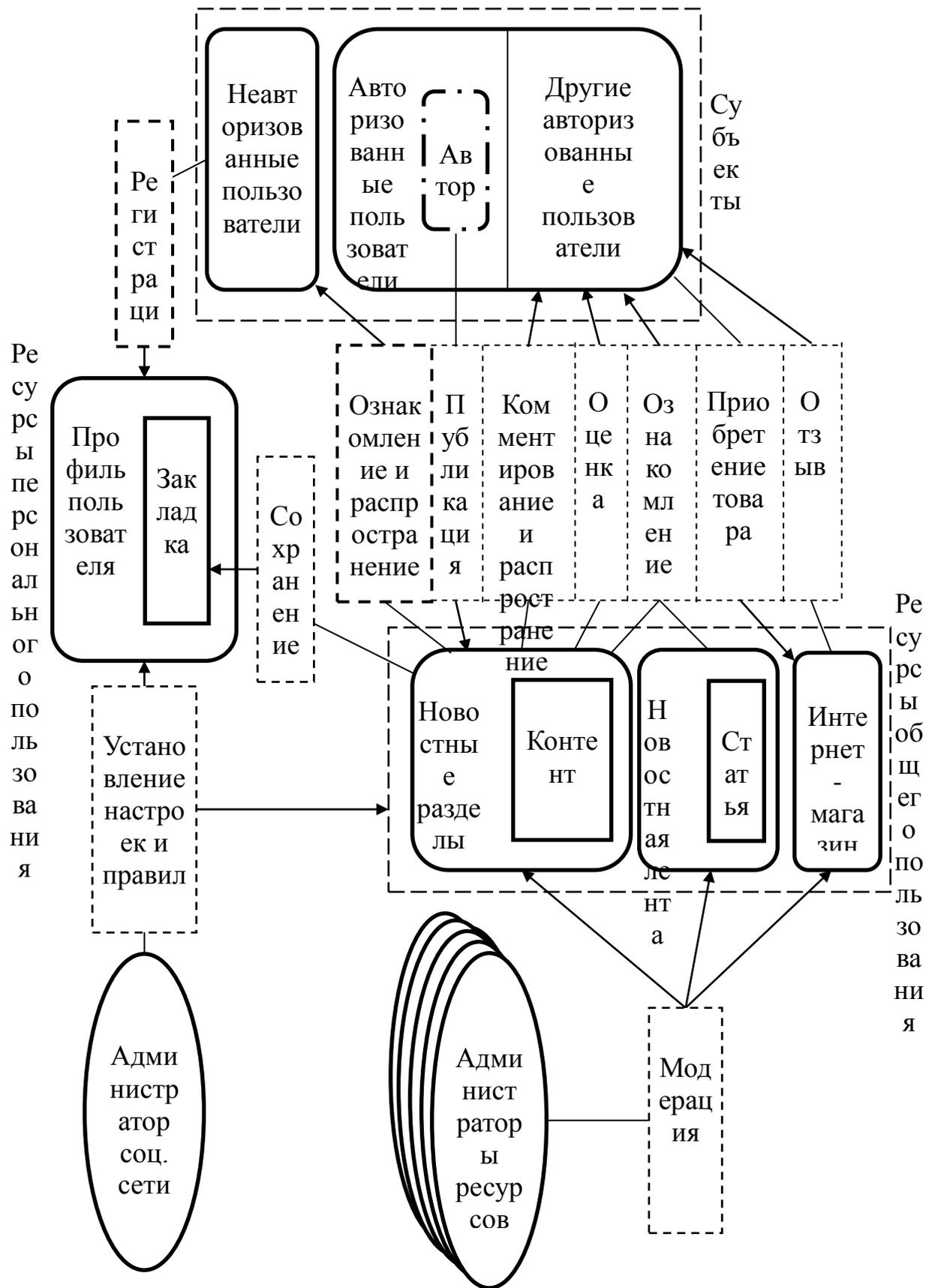


Рисунок 5 – Структурно-функциональная модель социальной сети Slashdot

Таким образом, в первой части данной статьи была дана подробная и всесторонняя классификация контента, циркулирующего в социальной сети Slashdot. Были рассмотрены и подробно описаны сетевые ресурсы данной социальной сети. Также в ходе анализа социальной сети Slashdot были проклассифицированы объекты и субъекты сети, взаимодействующие между собой. В ходе данного анализа было получено, что все субъекты с учетом проявления их активности в данной социальной сети можно разделить на активных и пассивных пользователей.

Также было установлено, что субъекты данной социальной сети способны обмениваться контентом посредством определенного набора действий (функций). Набор действий зависел от того авторизован пользователь или нет.

Всем актуальным для социальной сети Slashdot функциям пользователей была дана классификация по функциональному назначению.

На основе вышеприведенных данных была построена структурно-функциональная модель, которая учитывает все особенности данной социальной сети (рис. 5). На ней показаны функциональные связи между субъектами и объектам социальной сети Slashdot. А также отображены ресурсы личного и общего пользования.

Перейдем к статистическим данным, характеризующим количество узлов (пользователей) и ребер (дружеские связи, основанные на обмене закладками, комментировании статей и их публикации) в социальной сети Slashdot. Данные, изложенные на сайте Стенфордского Университета (Stanford Large Network Dataset Collection – Большая Стенфордская Коллекция Сетевых Данных) по социальной сети Slashdot, представлены в таблице 3 [5, 6, 7].

Статистика по социальной сети Slashdot собиралась в течение 9 лет, начиная с 2004 года.

Таблица 3 – Статистические данные социальной сети Slashdot

| Характеристические данные | Показатели |
|-------------------------------|-----------------------|
| Количество узлов | 62586 |
| Количество ребер | 515397 |
| Диаметр | 12 |
| Средняя длина пути | 3,99 |
| Коэффициент кластеризации (%) | 0,0006 |
| Распределение плотности сетей | $1,80282 \times 10^4$ |
| Диаметр сети | 4,3 |
| Средний диаметр сети | 3,9876 |

Используя приведенную статистику [5, 6, 7] необходимо всесторонне проанализировать социальную сеть Slashdot, провести репрезентативную выборку, а также доказать ее принадлежность степенному закону и доказать подобие полученной выборки и генеральной совокупности.

Статистические данные социальной сети Slashdot, показанные в таблице 3 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 самое ребро, и на третьем – вес ребра, отражающий силу связи между вершинами. Фрагмент такого предиката, представленном в текстовом формате, изобразим в таблице 4.

Таблица 4 – Фрагмент трехместного предиката социальной сети Slashdot

| Узел источник | Узел цель | Вес ребра | Узел источник | Узел цель | Вес ребра |
|---------------|-----------|-----------|---------------|-----------|-----------|
| 1 | 2 | 2 | 1 | 8 | 3 |
| 1 | 3 | 9 | 1 | 8 | 3 |
| 1 | 4 | 1 | 1 | 9 | 9 |
| 1 | 5 | 10 | 1 | 10 | 7 |
| 1 | 6 | 1 | 2 | 12 | 10 |
| 1 | 7 | 3 | 2 | 13 | 4 |

Формат данных в данном предикате соответствует формуле (1).

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

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 .

С учетом информации статистических сайтов [5, 6, 7] и описанию к ней, можно однозначно определить, что вес дуги $\delta(a_{ij})$ – есть величина, состоящая из следующих составляющих, взятых за определенный промежуток времени (1 месяц):

- 1) количество публикуемых статей пользователем;
- 2) количество комментариев и отзывов, написанных пользователем;
- 3) количество сохранений (создания закладок) пользователя.

В итоге, принимая во внимание вышеизложенное, получим общую формулу (2), характеризующую вес дуги для социальной сети Slashdot:

$$\delta(a_{ij}) = Ka_{ij} + Kr_{ij} + Ks_{ij}, \quad (2)$$

где Ka_{ij} – количество публикуемых статей пользователем;

Kr_{ij} – количество комментариев и отзывов, написанных пользователем;

Ks_{ij} – количество сохранений (создания закладок) пользователя.

Суть алгоритма преобразования исходных данных социальной сети Slashdot заключается в построении тройного предиката сначала в виде звездной матрицы, а затем в виде матрицы взвешенной центральности.

Рассмотрим данный алгоритм подробнее:

Шаг 1: Формирование звездной матрицы взвешенной связности;

Шаг 2: Определение диагональной матрицы степени вершин сети;

Шаг 3: Построение квадратной матрицы взвешенной центральности элементов сети;

Шаг 4: Построение диагональной матрицы удельного баланса трафика в вершинах сети.

На основе предложенного алгоритма преобразования исходных данных сети построим звездную матрицу социальной сети Slashdot (табл. 5) с использованием тройного предиката (табл. 4).

Таблица 5 – Звездная матрица социальной сети Slashdot

| | | | | | | | | | | | | | |
|-------|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-------|
| | 1 | 2 | 3 | ... | 11 | 12 | 13 | ... | 21 | 22 | 23 | ... | 62586 |
| 1 | - | 2 | 9 | ... | 10 | 0 | 0 | ... | 0 | 0 | 0 | ... | 0 |
| 2 | 0 | - | 0 | ... | 0 | 10 | 4 | ... | 3 | 0 | 0 | ... | 0 |
| 3 | 0 | 0 | - | ... | 0 | 0 | 0 | ... | 0 | 0 | 0 | ... | 0 |
| ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... |
| 11 | 0 | 0 | 0 | ... | - | 0 | 0 | ... | 0 | 0 | 0 | ... | 0 |
| 12 | 0 | 0 | 0 | ... | 0 | - | 0 | ... | 0 | 0 | 0 | ... | 0 |
| 13 | 0 | 0 | 0 | ... | 0 | 0 | - | ... | 0 | 0 | 0 | ... | 0 |
| ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... |
| 21 | 0 | 0 | 0 | ... | 0 | 0 | 0 | ... | - | 0 | 0 | ... | 0 |
| 22 | 0 | 0 | 0 | ... | 0 | 0 | 0 | ... | 0 | - | 0 | ... | 0 |
| 23 | 0 | 0 | 0 | ... | 0 | 0 | 0 | ... | 0 | 0 | - | ... | 0 |
| ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... |
| 62586 | 0 | 0 | 0 | ... | 0 | 0 | 0 | ... | 0 | 0 | 0 | ... | - |

В полученной матрице элементы строки соответствуют дугам, входящим в данную вершину, а элементы столбца – дугам, исходящим из вершины.

Второй шаг вышеизложенного алгоритма предлагает нам определить степень исхода для каждого узла социальной сети (табл. 6).

Полученная матрица (табл. 6) будет необходима для последующих процедур моделирования процессов диффузии контента в социальной сети.

Таблица 6 – Диагональная матрица степеней вершин сети Slashdot

| | | | | | | | | | | | | | |
|-------|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-------|
| | 1 | 2 | 3 | ... | 11 | 12 | 13 | ... | 21 | 22 | 23 | ... | 62586 |
| 1 | 10 | 0 | 0 | ... | 0 | 0 | 0 | ... | 0 | 0 | 0 | ... | 0 |
| 2 | 0 | 11 | 0 | ... | 0 | 0 | 0 | ... | 0 | 0 | 0 | ... | 0 |
| 3 | 0 | 0 | 3 | ... | 0 | 0 | 0 | ... | 0 | 0 | 0 | ... | 0 |
| ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... |
| 11 | 0 | 0 | 0 | ... | 17 | 0 | 0 | ... | 0 | 0 | 0 | ... | 0 |
| 12 | 0 | 0 | 0 | ... | 0 | 0 | 0 | ... | 0 | 0 | 0 | ... | 0 |
| 13 | 0 | 0 | 0 | ... | 0 | 0 | 16 | ... | 0 | 0 | 0 | ... | 0 |
| ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... |
| 21 | 0 | 0 | 0 | ... | 0 | 0 | 0 | ... | 0 | 0 | 0 | ... | 0 |
| 22 | 0 | 0 | 0 | ... | 0 | 0 | 0 | ... | 0 | 3 | 0 | ... | 0 |
| 23 | 0 | 0 | 0 | ... | 0 | 0 | 0 | ... | 0 | 0 | 0 | ... | 0 |
| ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... |
| 62586 | 0 | 0 | 0 | ... | 0 | 0 | 0 | ... | 0 | 0 | 0 | ... | 0 |

На следующем этапе вышеизложенного алгоритма необходимо построить квадратную матрицу взвешенной центральности элементов сети. Для определим удельный вес ее вершин и дуг с помощью нормировки их весов по суммарному трафику сети по следующей формуле:

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

где $\delta(a_{ij})$ – вес трафика в дуге a_{ij} .

Выражение (3) будет характеризовать степень взвешенной (по трафику) центральности дуги.

Далее, для определения взвешенной центральности вершины x_s , предлагается использовать пронормированную по суммарному трафику сети сумму.

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

С использованием (3) и (4) получим (5). Данное выражение будет характеризовать удельный вес трафика, по отношению ко всему трафику сети, проходящего через вершину x_s .

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

где $\delta(a_{si})$ – вес трафика в исходящей и узла a_{si} дуге;

$\delta(a_{js})$ – вес трафика во входящей в узел a_{js} дуге.

В итоге, следуя (5) получим матрицу взвешенной центральности элементов исходной сети следующего вида (табл. 7).

Таблица 7 – Матрица взвешенной центральности социальной сети Slashdot

| | | | | | | |
|-------|--------|----------|-----|---------|-----|-----------|
| | 1 | 2 | ... | 11 | ... | 62586 |
| 1 | 0,0007 | 0,00002 | ... | 0 | ... | 0 |
| 2 | 0 | 0,000114 | ... | 0,00012 | ... | 0 |
| ... | ... | ... | ... | ... | ... | ... |
| 22 | 0 | 0 | ... | 0 | ... | 0 |
| ... | ... | ... | ... | ... | ... | ... |
| 62586 | 0 | 0 | ... | 0 | ... | 0,0000184 |

На заключительном этапе будет учитываться разность исходящего и входящего для вершины трафиков:

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

где $\delta(a_{si})$ – вес трафика в исходящей из узла a_{si} дуге;

$\delta(a_{js})$ – вес трафика во входящей в узел a_{js} дуге.

Данный шаг необходим для построения диагональной матрицы удельного баланса трафика.

После применения (6), получим диагональную матрицу удельного баланса трафика в вершинах сети (табл. 8).

Таблица 8 – Матрица удельного баланса трафика в вершинах социальной сети Slashdot

| | | | | | | |
|-------|----------|----------|-----|----------|-----|----------|
| | 1 | 2 | ... | 11 | ... | 62586 |
| 1 | -0,00092 | 0 | ... | 0 | ... | 0 |
| 2 | 0 | -0,00023 | ... | 0 | ... | 0 |
| ... | ... | ... | ... | ... | ... | ... |
| 11 | 0 | 0 | ... | -0,00038 | ... | 0 |
| ... | ... | ... | ... | ... | ... | ... |
| 62586 | 0 | 0 | ... | 0 | ... | -0,00031 |

Полученная удельная величина характеризует 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.

Анализ социальной сети затрудняет ее размер (десятки тысяч вершин). Для того чтобы, в полной мере проанализировать данную сеть необходимо преобразовать всю генеральную совокупность тройного предиката (табл. 4) в некоторую выборку, которая описывает и показывает все основные свойства сети.

Данный подход заключается в отборе вершин по их степени, то есть количеству инцидентным им дуг. Однако такой подход не учитывает тот факт, что каждая дуга имеет свой вес.

Таким образом, отсутствует взвешенная оценка с точки зрения трафика. На основе этого предлагается использовать следующий алгоритм:

Шаг 1: Вычленение из квадратной матрицы взвешенной центральности элементов сети ее диагонали;

Шаг 2: Пересортировка последовательности расположения вершин в диагонали по убыванию;

Шаг 3: Удаление из усеченной матрицы все ее элементы, относящиеся к исключенным выборкой вершинам;

Шаг 4: Виртуализация выборки;

Шаг 5: Формирование итоговой матрицы.

Суть данного алгоритма состоит в том, что требуется из полученных квадратных матриц взвешенности, получить усеченное количество узлов, которые способны с сохранением подобия описать исходную сетевую структуру.

Поэтому в предложенном алгоритме уместно использование квадратной матрицы взвешенной центральности элементов сети, а точнее – ее диагонали, где определен показатель степени взвешенной центральности каждой вершины (табл. 9).

Таблица 9 – Показатели степени взвешенной центральности вершин

| Номер узла | Показатель | Номер узла | Показатель |
|------------|--------------|------------|--------------|
| 1 | 0,0000767234 | 22 | 0,0000049103 |
| 2 | 0,0001141644 | 23 | 0,0000257790 |
| 3 | 0,0000227101 | 24 | 0,0000079792 |
| 4 | 0,0000939094 | 25 | 0,0000098206 |
| ... | ... | ... | ... |
| 20 | 0,0000626063 | 62585 | 0,0000030689 |
| 21 | 0,0000018414 | 62586 | 0,0000018414 |

На втором этапе данного алгоритма необходимо проранжировать показатели степени узлов матрицы по убыванию (табл. 10).

Таблица 10 – Проранжированные показатели степени взвешенной центральности вершин

| Номер узла | Показатель | Номер узла | Показатель |
|------------|-------------|------------|-------------|
| 9788 | 0,006211523 | 27219 | 0,003056659 |
| 17325 | 0,005376773 | 830 | 0,003019831 |
| 28802 | 0,004443817 | 32727 | 0,003019831 |
| 50445 | 0,004222854 | 26527 | 0,003007556 |
| ... | ... | ... | ... |
| 33581 | 0,003661513 | 5874 | 0,002461389 |
| 5989 | 0,003661513 | 4211 | 0,002461389 |
| 62223 | 0,003661513 | 54669 | 0,002461389 |

Продолжение таблицы 10

| | | | |
|-------|-------------|-------|-------------|
| ... | ... | ... | ... |
| 36548 | 0,003105775 | 2658 | 0,000012389 |
| 1191 | 0,003105762 | 62572 | 0,000012276 |
| 26459 | 0,003105762 | 62580 | 0,000012276 |

На следующем этапе просуммируем полученные значения для получения необходимого результата (точности модели). То есть до значения 0.95, так как допустима 5% потеря трафика. Полученная сумма должна быть не меньше 0,95 [8].

Этим критерием и ограничится репрезентативная (с точки зрения трафика) выборка:

$$\delta(\bar{x}_s) = 0,006211523 + 0,005376773 + 0,004443817 + \dots + 0,003105762 + 0,003056659 + \dots + 0,0001853636 = 0,9501.$$

При данном суммировании имеем выборку из 382 узлов, с помощью которых можно описать данную сеть с 5% погрешностью.

Целью данной визуализации является проверка выборки на отсутствие несвязных кластеров.

Далее обнулим столбцы и строки, на пересечении которых стояла исключенная вершина. На основе данного усечения построим следующую матрицу (табл. 11).

Таблица 11 – Проранжированная усеченная матрица социальной сети Slashdot

| | | | | | | | |
|-------|------|--------|--------|-----|--------|-----|-------|
| | 9788 | 17325 | 28802 | ... | 44323 | ... | 61758 |
| 9788 | 0 | 0 | 0 | ... | 0 | ... | 0 |
| 17325 | 0 | 0 | 0,0022 | ... | 0 | ... | 0 |
| 28802 | 0 | 0,0026 | 0 | ... | 0,0017 | ... | 0 |
| ... | ... | ... | ... | ... | ... | ... | ... |
| 44323 | 0 | 0 | 0 | ... | 0 | ... | 0 |
| ... | ... | ... | ... | ... | ... | ... | ... |
| 61758 | 0 | 0 | 0 | ... | 0 | ... | 0 |

На основе полученной матрицы (табл. 11), с помощью программного обеспечения Gephi [9] построим усеченный граф социальной сети (рис. 6):

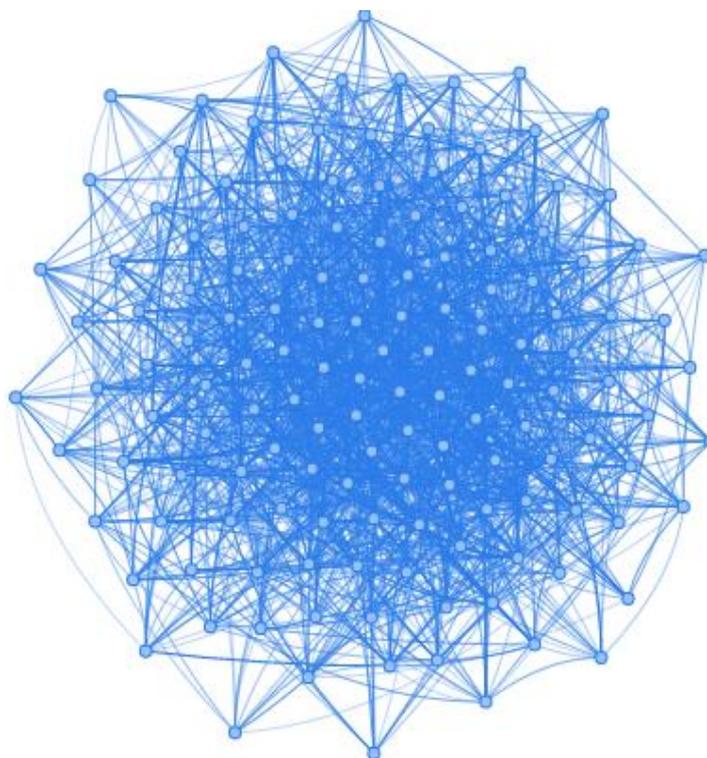


Рис. 6. Выборка социальной сети Slashdot

Стоит отметить, что у данной социальной сети явная кластерность не прослеживается. Это напрямую зависит от возможностей авторизованных пользователей – любой авторизованный пользователь может прочесть любую статью, прокомментировать ее и сохранить к себе в закладки. При этом в данной сети нет групп, в которые пользователи могут объединять для обсуждения только им интересных вопросов. В социальной сети Slashdot все ресурсы и контент общедоступны [4].

Последний этап алгоритма состоит в формировании итоговой матрицы в виде, удобном для последующего моделирования информационной диффузии (табл. 12). Отсюда определяется ожидаемое количество вирусных контентом элементов.

Таблица 12 – Модифицированная матрицей смежности социальной сети Slashdot

| | | | | | | | | |
|-------|------|-----|-------|-------|-----|-------|-----|-------|
| | 9788 | ... | 44323 | 13596 | ... | 26459 | ... | 61758 |
| 9788 | 0 | ... | 0 | 0 | ... | 0 | ... | 0 |
| ... | ... | ... | ... | ... | ... | ... | ... | ... |
| 44323 | 0 | ... | 7 | 0 | ... | 0 | ... | 0 |
| 13596 | -1 | ... | 0 | 5 | ... | 0 | ... | 0 |
| ... | ... | ... | ... | ... | ... | ... | ... | ... |
| 26459 | 0 | ... | 0 | 0 | ... | 11 | ... | 0 |
| ... | ... | ... | ... | ... | ... | ... | ... | ... |
| 61758 | 0 | ... | 0 | 0 | ... | 0 | ... | 28 |

В данной матрице недиагональные элементы равны ± 1 , так как интересуется лишь факт смежности вершин истоков (+1) и стоков (-1), а диагональные элементы имеют вес k , равный сумме элементов (+1) столбца.

Таким образом, после применения предложенных алгоритмов преобразования исходных данных сети и нахождения репрезентативной выборки была получена визуальная модель усеченной социальной сети Slashdot и соответствующие матрицы, способные в полной мере охарактеризовать распространение контента в данной социальной сети.

Для того чтобы с уверенностью можно было пользоваться усеченной сетью, необходимо доказать подобие основной сети и ее выборки.

Возможным способом доказать подобие основной сети и полученной в результате применения алгоритма (на основе доказательства о принадлежности полученной выборки сети к степенному закону распределения) является критерий Пирсона [10, 11].

Посчитаем критерий Пирсона для выборки сети Slashdot.

$$\chi_{\text{набл}}^2 = \sum_{i=1}^k \frac{(n - np_i)^2}{np_i}, \quad (7)$$

где n – объем выборки;

p_i – вероятность попадания случайной величины X в заданный интервал степеней.

Следуя выражению (7), рассчитаем значение χ^2 . В данном случае $\chi^2=20.41428$. Следовательно, $n = 20$ (выборка) – 1 (количество параметров в плотности степенного распределения) = 19 (число степеней свободы).

Далее выберем уровень значимости $\alpha = 0.05$, тогда $p = 1 - \alpha = 0.95$ [12]. Табличное значение χ^2 равно 30.1. Получается, что табличное значение χ^2 больше вычисленного. Следовательно, гипотеза верна, и приведенное степенное распределение согласуется с результатами эксперимента, что и требовалось доказать [8].

На следующем этапе нашего исследования необходимо провести расчет среднеквадратического отклонения (СКО) для исходной сети и полученной репрезентативной выборки по формулам (8) и (9):

$$\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i; \quad (8)$$

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

где: n – общее число вершин;

X_i – i -ый элемент выборки;

\bar{X} – среднее арифметическое выборки;

σ – среднеквадратическое отклонение.

Тогда получим: $\sigma_1 = 0.289$ – СКО генеральной совокупности; $\sigma_2 = 0.276$ – СКО репрезентативной выборки.

Отклонение между эталоном (генеральной совокупностью) и полученной выборкой получилось равным 5%. Это доказывает, что репрезентативная выборка подобна генеральной совокупности с заданной точностью [8].

Построим гистограммы распределения вершин для генеральной совокупности и полученной выборки (по горизонтальной оси – степени вершин, по вертикальной – количество вершин данной степени), иллюстрирующих полученные ранее результаты (рис. 7).

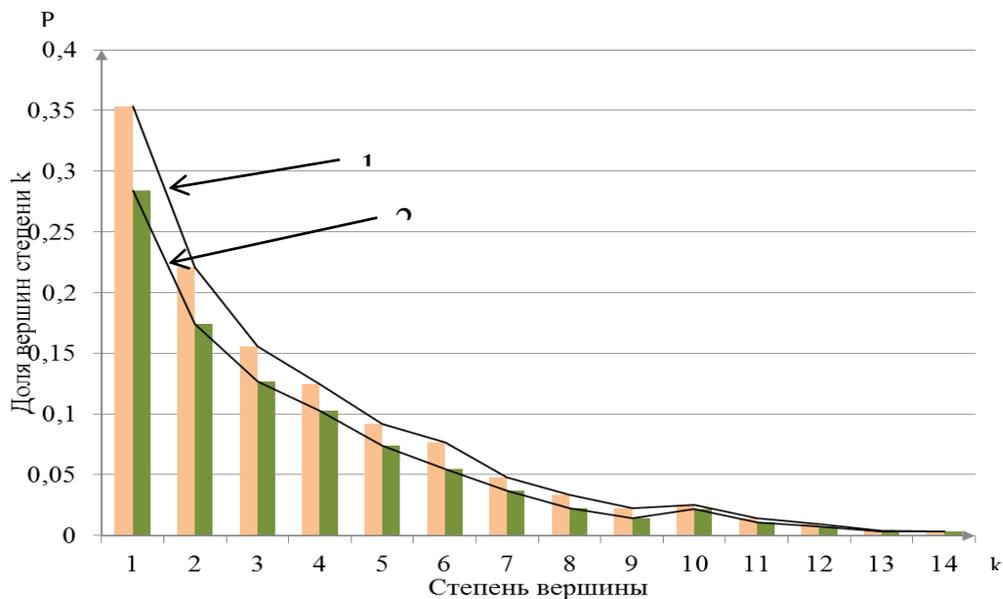


Рис. 7. Гистограмма распределения вершин

где: 1 – огибающая распределения степеней вершин генеральной совокупности;

2 – огибающая распределения степеней вершин репрезентативной выборки;

k – степень вершины;

$P(k)$ – вероятность попадания вершины со степенью k в заданный интервал.

Таким образом, опираясь на передовые исследования отечественных и зарубежных ученых [13-19], были выполнены алгоритмы преобразования исходных данных сети и нахождения репрезентативной выборки, получена визуальная модель исследуемой сети, а также вычислены соответствующие матрицы, позволяющие провести дальнейший анализ с целью получения полной картины распространения контента в социальной сети.

Далее по критерию Пирсона была проверена генеральная совокупность социальной сети Slashdot и репрезентативная выборка, полученная в ходе анализа сети, на соответствие степенному закону распределения (с заданной погрешностью 5%). Для данных совокупностей было получено среднеквадратическое отклонение и сопоставлено показано сравнение исходной генеральной совокупности и полученной выборки, состоящей из 382 узлов на соответствующем графике (рис. 7).

Литература:

1 Абрамов К.Г. Распространение нежелательной информации в социальных сетях Интернета / К.Г. Абрамов //Медиа индустрия и информационная безопасность. – № 3. – С. 45-48.

2 Тищенко В. И. Социальные сети и виртуальные сетевые сообщества / Тищенко В.И., Верченков Л.Н., Ефременко Д.В. – М., 2013. – 360 с.

3 Назарчук А.В., О сетевых исследованиях в социальных науках / МГУ им. М. В. Ломоносова – М.: Типография МГУ, 2008. – 73 с.

4 Социальная сеть Slashdot. – Электрон. Дан. – Режим доступа: <http://slashdot.org> (дата обращения: 16.07.2016)

5 Статистика социальной сети Slashdot. – Электрон. Дан. – Режим доступа: <http://konect.uni-koblenz.de/networks/slashdot>. (дата обращения: 12.09.2016)

6 Большая Стенфордская Коллекция Сетевых Данных – Электрон. Дан. – Режим доступа: <https://snap.stanford.edu/data/>

7 Статистические данные посещения социальной сети. – Электрон. Дан. – Режим доступа: <http://www.alexa.com/>

8 Губанов Д.А., Новиков Д.А., Чхаратишвили А.Г. Социальные сети: модели информационного влияния, управления и противоборства /

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

9 Средство визуализации данных. – Электрон. Дан. – Режим доступа: <https://gephi.org/> (дата обращения: 15.09.2016)

10 Гмурман В.Е., Теория вероятностей и математическая статистика. Учебное пособие. Высшее образование. – Москва, 2006 – С. 243.

11 Статистический анализ данных, моделирование и исследование вероятностных закономерностей. Компьютерный подход : монография / Б. Ю. Лемешко, С. Б. Лемешко, С. Н. Постовалов, Е. В. Чимитова. – Новосибирск : Изд-во НГТУ, 2011. – 888 с.

12 George Casella, Roger L. Berger. Hypothesis Testing // Statistical Inference. – Second Edition. – Pacific Grove, CA: Duxbury, 2002. – С. 397. – 660 с.

13 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.

14 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.

15 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.

16 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.

17 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.

18 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.

19 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.

Воронежский научно-образовательный центр
управления информационными рисками

**СТРУКТУРНО – ФУНКЦИОНАЛЬНАЯ СПЕЦИФИКА
РАСПРОСТРАНЕНИЯ КОНТЕНТА В СОЦИАЛЬНОЙ СЕТИ
RATE&GOODS**

А.В. Паринов, Д.М. Баранов

Социальная сеть отзывов и обзоров Rate&Goods (Ratengoods.com) – это совместный проект инвестиционного фонда ItRuStore и Ассоциации ГС1РУС [1].

Для постинга в данной социальной сети выделяют три основных формата контента [2, 3]: текстовый (статья, обсуждение, интервью, ссылка, списки товаров), графический (картинка, рисунок, фотография, карта) и гибридный, получаемый при совмещении двух предыдущих (уведомления, события, пост) (рис. 1).

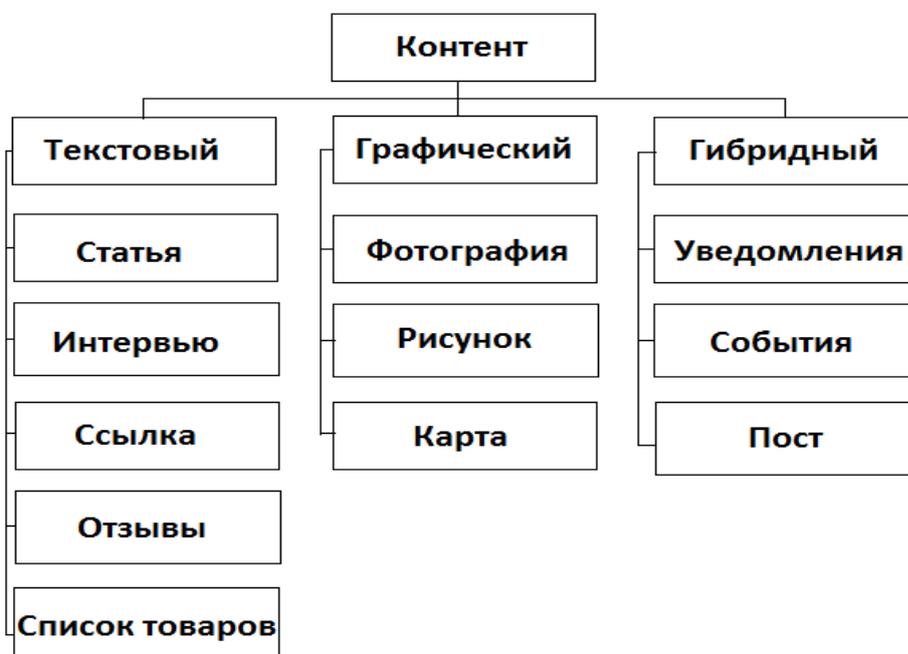


Рис. 1. Классификация контента сети Rate&Goods по форматам

Структурно-функциональная специфика социальной сети “Rate&Goods” является весьма уникальной. Ratengoods подключено к базе

данных ГС1РУС, содержащей данные о всех отечественных товарах и производителях, а также к глобальному информационному сервису GS1 GEPiR, включающему сведения о всех зарубежных производителях [1, 4].

Рассмотрим структуру размещения контента в социальной сети отзывов и закладок Rate&Goods (рис. 2) [1, 3]:

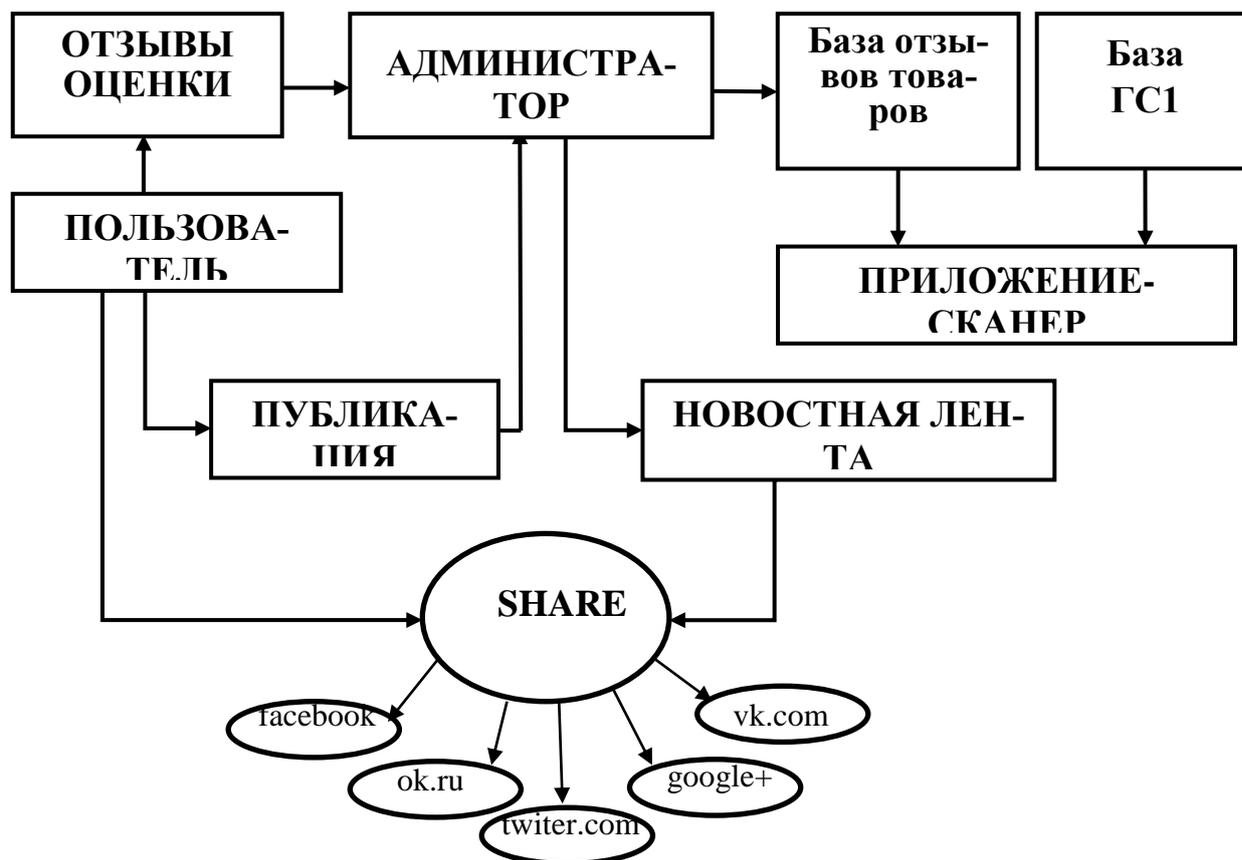


Рис. 2. Обобщенная схема размещения контента в сети Rate&Goods

1. В сети Rate&Goods одним из основных узлов является приложение-сканер, которое помогает определить рассматриваемый товар.

2. При сканировании товара, посылается запрос в базу ГС1 (ассоциация автоматической идентификации) и базу товара (дополнительные отзывы и описания), которую пополняют сами пользователи.

3. Далее, отсканировав товар, можно просмотреть все комментарии, отзывы, касающиеся его на отдельной странице, а дальнейшем, оставить оценку, комментарий, репост или отзыв на данной странице.

4. Все записи, учитывающиеся в базе товаров, перед размещением проверяются администрацией Rate&Goods.

5. При положительной проверке, записываются в базе товаров.

6. Также в данной социальной сети можно вести свои блоги и посты.

7. В приложении они отдельно размещаются заголовками с хештегами. Однако размещение подобных записей также согласуется с администрацией.

8. В сети Ratengoods.com своими публикациями возможно делиться (SHARE) в таких социальных сетях как вконтакте, одноклассники, twitter, google+ и facebook и других (рис. 2).

Ресурсы Rate&Goods делятся на персональные (закладки и данные аккаунта пользователя) и коллективные (лента, история профиля, интерактивная карта и справка) (рис. 3).

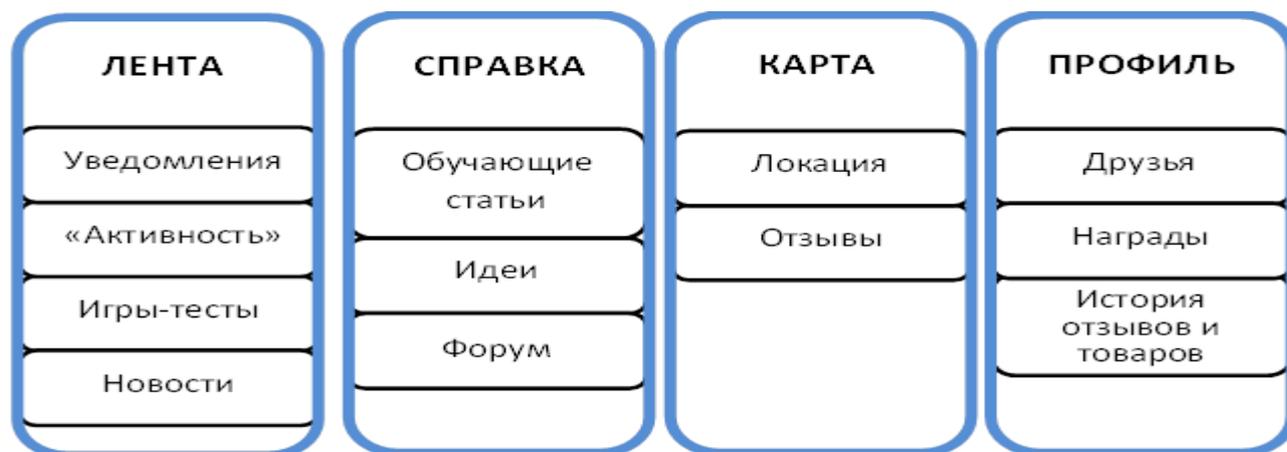


Рис. 3. Классификация объектов, создаваемых в ресурсах сети Rate&Goods

Доступ к персональным ресурсам пользователя ограничен. Допуск к таким данным имеет только пользователь и администратор. Ресурсы коллективного пользования увидеть и прочитать может и зарегистрированный, и незарегистрированный пользователь в сети, однако размещать контент разрешено только после авторизации в Rate&Goods. Гостю сети также

разрешено использовать приложение для сканирования штрих-кода товара и просмотра отзывов о нем [5, 6].

Размещение контента делится на публикацию в ресурсе (обзор, отзыв, комментарий, репост) и на добавление товара в приложении. Можно заметить, что действия по размещению и реагированию тесно связаны (рис. 4).



Рис. 4. Классификация действий субъектов в соцсети Rate&Goods

Реагируют на обновление какого-либо контента в соцсети как администраторы (управленческие), так и пользователи (пользовательские). Администрация контролирует публикацию новых обзоров товаров, а также не допускает нарушения правил данной социальной сети.

В связи с этим, можно утверждать, что модератор и системный администратор являются постоянными активными субъектами сети Rate&Goods [7]. Также к ним можно отнести всех пользователей, оценивающих, комментирующих и публикующих записи в соцсети. Пассивными

субъектами являются гости или просто зарегистрированные читатели, не распространяющие контент (рис. 5, 6).

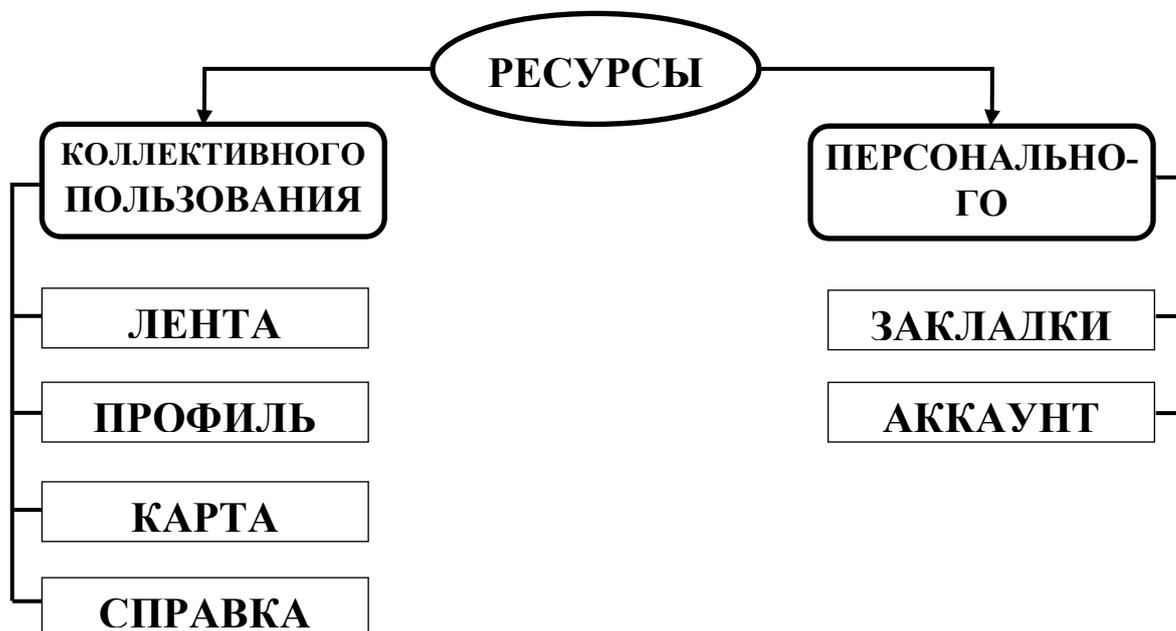


Рис. 5. Классификация субъектов, функционирующих в соцсети Rate&Goods

Все посетители сети “Rate&Goods” делятся на пассивных пользователей, не воздействующих на перемещение контента, и активных, которые влияют не только на распространение информации, но и самостоятельно размещают её в интернет-ресурсе (рис. 6). По приоритету они классифицируются на [6, 7]:

1. VIP-пользователей;
2. Экспертов;
3. Генерирующих контент;
4. Репостеров;
5. Читателей;
6. Гостей.

Пассивные юзеры, гости (6) сети и читатели (5), отличаются друг от друга фактом регистрации последнего в сети. При проявлении активных действий (написании чего-либо или репосте) читатель автоматически пе-

переходит в разряд генерирующего контент (3) или репостера (4). Зарегистрируясь в ресурсе и выполняя условия, описанные ранее, гость так же перемещается в 3 или 4 классы [7].

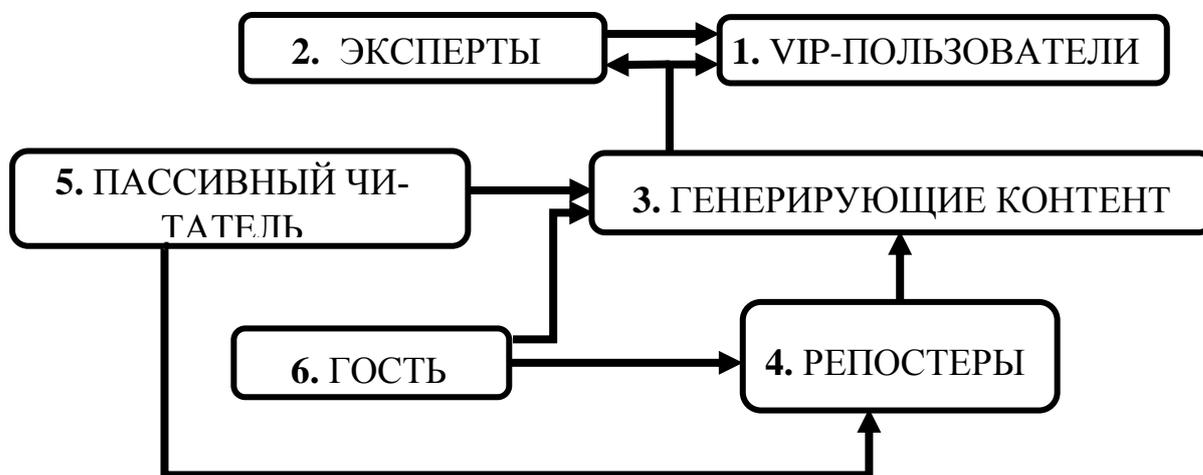


Рис. 6. Схема переходов пользователей ресурса по приоритету

Для того чтобы стать “экспертом” сети “Rate&Goods” необходимо находиться в 3 категории пользователей, иметь достаточное количество личных постов, а также отправить заявку администрации. Если все требования соблюдены и обращение об “эксперте” подтверждено, то юзера сети переводят во 2 класс [1, 7].

Высший разряд интернет-ресурса считается VIP-пользователь. Ими могут стать либо публичные и знаменитые люди, либо эксперты, выделяющиеся среди остальных большей активностью и авторитетностью среди других пользователей сети. Добавление в 1 класс осуществляется при подтверждении самим юзером обращения администрации о включении в VIP класс [3, 7].

Все перечисленные активные пользователи имеют разграничения по размещению контента в исследуемом интернет-ресурсе. Подробнее их вза-

взаимодействия с объектами сети через ресурсы проиллюстрированы на рисунке 7.

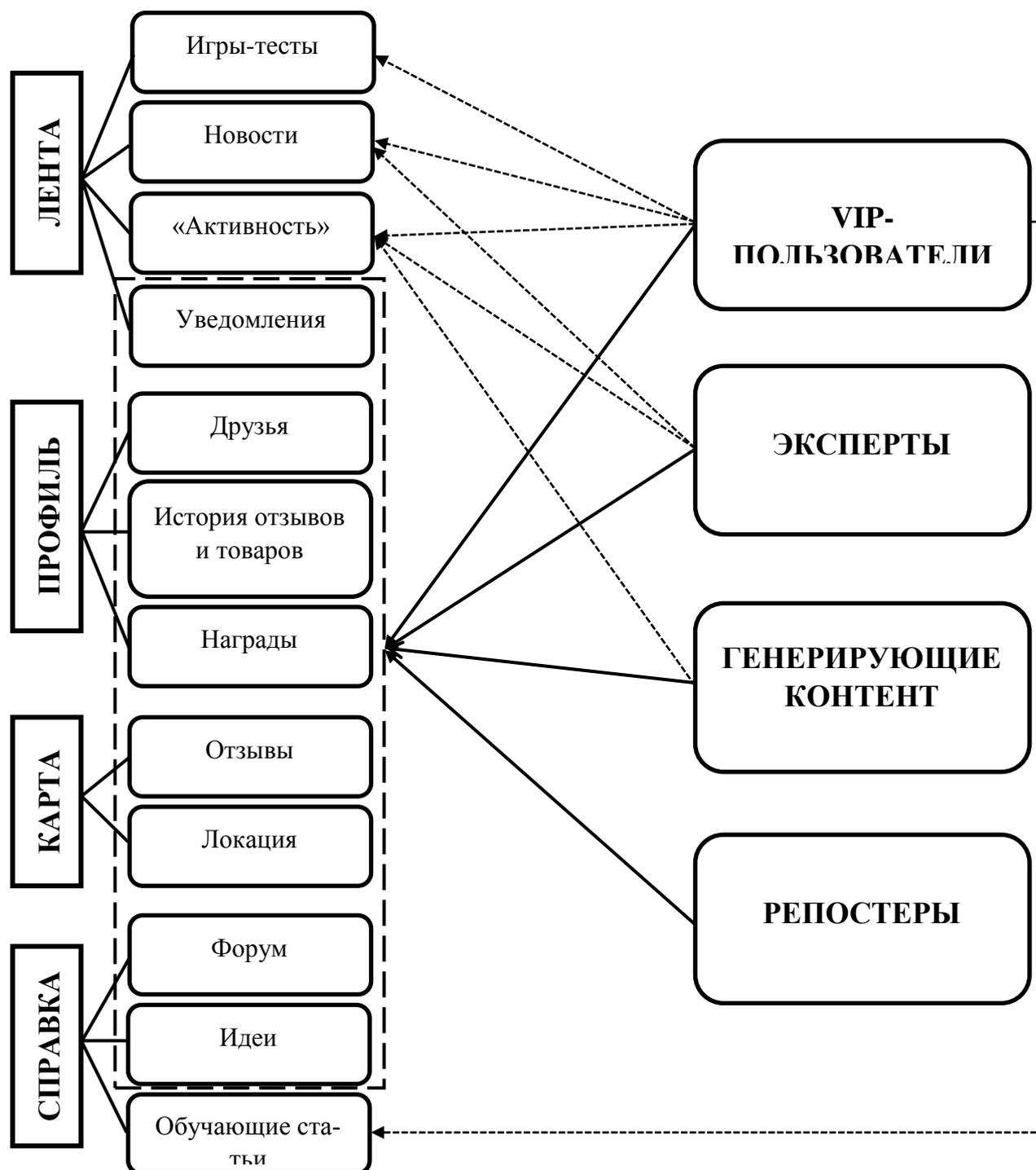


Рис. 7. Взаимодействие активных пользователей и объектов сети

Отсюда представляется возможным построить обобщенную структурно-функциональную схему социальной сети Rate&Goods (рис. 8).

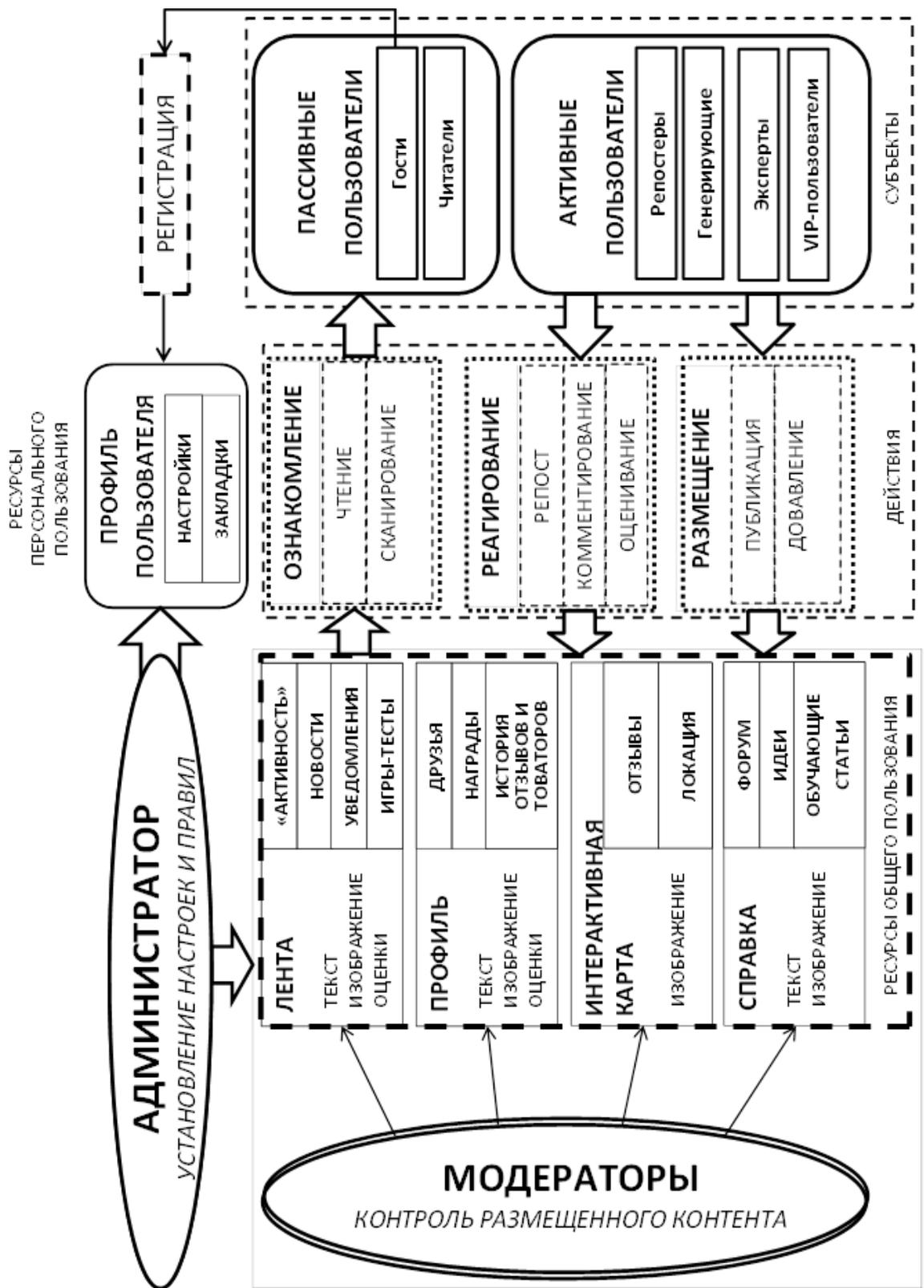


Рис. 8. Обобщенная структурно-функциональная схема социальной сети Rate&Goods

По данным администрации исследуемой социальной сети [7], динамика популяции зарегистрированных пользователей сети «Rate&Goods» за последние 6 месяцев положительна. Для сравнения на 7.05.2016 равно 1574, а на 7.10.2016 – 8371 человек [5, 7].

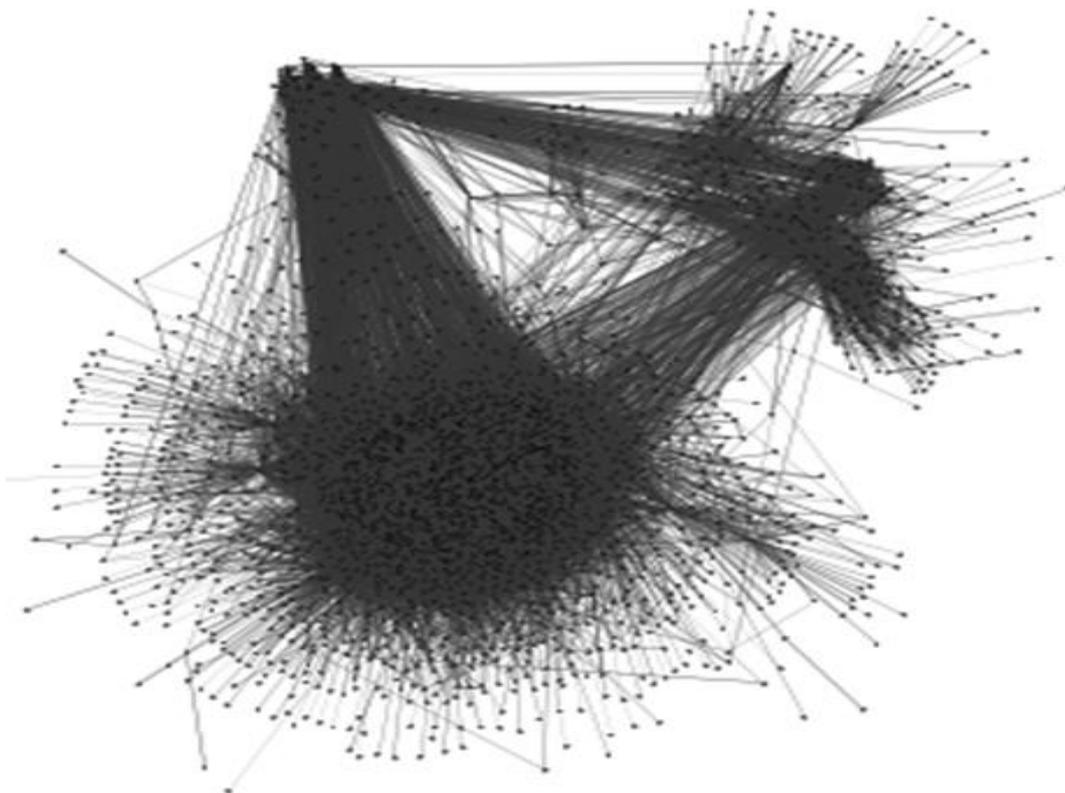


Рис. 9. Графическая модель социальной сети Rate&Goods на 7.10.2016

Пока в сети много одноразовых пользователей – гостей. Их количество в настоящее время около 500.000 [7]. Вследствие этого, при анализе структуры сети, эти вершины (люди) могут не учитываться.

Используя программное обеспечение «Gephi» [8], месячные статистические данные (7.10.2016), предоставленные администрацией исследуемой соцсети, [7] и трехместный предикат сети Rate&Goods, графически представим исследуемую социальную сеть без узлов администрации в виде графа (рис. 9), в котором вершины являются пользователями, а взвешенные дуги обозначают количество репостов обзоров товаров Rate&Goods за месяц [7].

Отсюда сеть можно представить в виде звездной матрицы (табл. 1). Произведя репрезентативную выборку, можно получить распределение сети с меньшим количеством вершин и при этом иметь подобную структуру, что и при генеральной совокупности элементов.

Таблица 1 – Звездная матрица для социальной сети Rate&Goods

| № узла | 1 | 2 | 3 | ... | 18 | 19 | ... | 101 | 101 | 102 | ... | 837 |
|--------|---|---|---|-----|----|----|-----|-----|-----|-----|-----|-----|
| 1 | - | 5 | 4 | ... | 0 | 0 | ... | 2 | 0 | 0 | ... | 0 |
| 2 | 1 | - | 0 | ... | 0 | 3 | ... | 0 | 5 | 0 | ... | 0 |
| 3 | 0 | 0 | - | ... | 0 | 0 | ... | 0 | 3 | 0 | ... | 0 |
| ⋮ | ⋮ | ⋮ | ⋮ | ⋮ | ⋮ | ⋮ | ⋮ | ⋮ | ⋮ | ⋮ | ⋮ | ⋮ |
| 18 | 0 | 0 | 0 | ... | - | 0 | ... | 0 | 0 | 2 | ... | 0 |
| 19 | 0 | 3 | 0 | ... | 0 | - | ... | 0 | 4 | 0 | ... | 0 |
| ⋮ | ⋮ | ⋮ | ⋮ | ⋮ | ⋮ | ⋮ | ⋮ | ⋮ | ⋮ | ⋮ | ⋮ | ⋮ |
| 1018 | 0 | 0 | 0 | ... | 0 | 0 | ... | - | 0 | 0 | ... | 0 |
| 1019 | 0 | 0 | 0 | ... | 0 | 0 | ... | 0 | - | 2 | ... | 0 |
| 1020 | 0 | 0 | 0 | ... | 0 | 0 | ... | 0 | 2 | - | ... | 1 |
| ⋮ | ⋮ | ⋮ | ⋮ | ⋮ | ⋮ | ⋮ | ⋮ | ⋮ | ⋮ | ⋮ | ⋮ | ⋮ |
| 8371 | 0 | 0 | 0 | ... | 0 | 0 | ... | 0 | 0 | 0 | ... | - |

Формируя выборку, необходимо очень внимательно следить за тем, чтобы не только отобрать из всей совокупности элементов сети достаточное количество объектов, но и взять такую группу, которая, как представляется, будет действительно репрезентативной с точки зрения трафика [9].

Из полученной матрицы (табл. 1) определяются веса узлов $\delta(x_s)$ по сумме весов всех входящих $\delta(a_{si})$ и исходящих $\delta(a_{js})$ дуг сети. Нормирование полученных параметров вершин помогает построить квадратную матрицу взвешенной центральности элементов сети, диагональные элементы которой определяются следующим образом:

$$\delta(\bar{x}_s) = \frac{\sum_i \delta(a_{si}) + \sum_j \delta(a_{js})}{\sum_{\substack{i,j \\ i \neq j}} \delta(a_{ij})}, \quad (1)$$

где значение $\delta(\bar{x}_s)$ характеризует удельный вес трафика, проходящего через вершину x_s , по отношению ко всему трафику сети $\sum_{i \neq j} \delta(a_{ij})$ [11, 12].

Следуя выражению (1), можно получить квадратную матрицу взвешенной центральности элементов исходной сети. Для наглядного обзора, сети выбраны ключевые вершины графа (табл. 2).

Таблица 2 – Матрица взвешенной центральности для сети Rate&Goods

| | | | | | | | | | |
|--------|---------------------|-----|----------------------|-----|---------|-----|---------|-----|----------|
| № узла | 1 | ... | 46 | ... | 69 | ... | 74 | ... | 8371 |
| 1 | $0,7 \cdot 10^{-5}$ | ... | $0,12 \cdot 10^{-6}$ | ... | 0 | ... | 0 | ... | 0 |
| ... | ... | ... | ... | ... | ... | ... | ... | ... | ... |
| 46 | 0 | ... | 0,01058 | ... | 0,00045 | ... | 0 | ... | 0 |
| ... | ... | ... | ... | ... | ... | ... | ... | ... | ... |
| 69 | 0 | ... | 0 | ... | 0,00948 | ... | 0 | ... | 0 |
| ... | ... | ... | ... | ... | ... | ... | ... | ... | ... |
| 74 | 0 | ... | 0,00127 | ... | 0 | ... | 0,00746 | ... | 0 |
| ... | ... | ... | ... | ... | ... | ... | ... | ... | ... |
| 8371 | 0 | ... | 0 | ... | 0 | ... | 0 | ... | 0,000018 |

Следующий этап преобразования исходных данных сети заключается в нахождении и построении диагональной матрицы удельного баланса трафика в вершинах социальной сети Rate&Goods (табл. 3).

В данной таблице отражена значимость узла сети, определяемая по формуле:

$$\delta(\bar{B}_s) = \frac{[\sum_i \delta(a_{si}) - \sum_j \delta(a_{js})]}{\sum_{i \neq j} \delta(a_{ij})}, \quad (2)$$

где: $\delta(a_{si})$ – вес трафика в исходящей и узла a_{si} дуге;

$\delta(a_{js})$ – вес трафика во входящей в узел a_{js} дуге.

Таблица 3 – Диагональная матрица удельного баланса трафика в вершинах сети Rate&Goods

| | | | | | | | | | | | |
|--------|----------|-----|---------|-----|---------|-----|---------|-----|-----------|-----|---------|
| № узла | 2 | ... | 46 | ... | 69 | ... | 74 | ... | 150 | ... | 8371 |
| 2 | -0,00036 | ... | 0 | ... | 0 | ... | 0 | ... | 0 | ... | 0 |
| ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... |
| 46 | 0 | ... | 0,00019 | ... | 0 | ... | 0 | ... | 0 | ... | 0 |
| ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... |
| 69 | 0 | ... | 0 | ... | 0,00012 | ... | 0 | ... | 0 | ... | 0 |
| ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... |
| 74 | 0 | ... | 0 | ... | 0 | ... | 0,00025 | ... | 0 | ... | 0 |
| ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... |
| 150 | 0 | ... | 0 | ... | 0 | ... | 0 | ... | -0,000014 | ... | 0 |
| ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... |
| 8371 | 0 | ... | 0 | ... | 0 | ... | 0 | ... | 0 | ... | 0,00023 |

Полученная удельная величина при помощи выражения (2), характеризует роль вершины (субъекта) сети: при $\bar{B}_s > 0$ – генератор контента; при $\bar{B}_s < 0$ – аккумулятор контента. Чем больше модуль \bar{B}_s , тем в большей степени проявляется его ролевая функция [11, 12].

Однако, необходимо помнить, что данная социальная сеть постоянно развивается [7] и созданная база данных (табл. 1) требует систематической актуализации.

Стандартный метод определения репрезентативной выборки заключается в отборе вершин по их степени, т.е. количеству инцидентным им дуг. Вместе с тем, такой критерий явно не учитывает тот факт, что каждая дуга имеет свой вес, т.е. отсутствует взвешенная оценка с точки зрения трафика.

В связи с этим, уместно использование квадратной матрицы взвешенной центральности элементов сети (табл. 2), а точнее – ее диагонали, где определена степень взвешенной центральности каждой вершины.

Произведем ранжирование вершин в порядке убывания степеней взвешенной центральности вершин (табл. 4, 5).

Таблица 4 – Вершины сети, ранжированные по убыванию

| id | степень | id | степень |
|------|-------------|------|-------------|
| 46 | 0,010575063 | 133 | 0,007469969 |
| 880 | 0,009741124 | 80 | 0,007434483 |
| 69 | 0,009386256 | 62 | 0,006724747 |
| 74 | 0,009333026 | 196 | 0,00663603 |
| 165 | 0,009262052 | 5015 | 0,006440852 |
| 150 | 0,008853954 | 1905 | 0,006263418 |
| 2733 | 0,008818467 | 54 | 0,006227932 |
| 174 | 0,008818467 | 136 | 0,006227932 |
| 159 | 0,008747494 | 157 | 0,006156958 |
| 57 | 0,008374882 | 51 | 0,006121471 |
| 3127 | 0,008357139 | 802 | 0,006121471 |
| 60 | 0,008215192 | 106 | 0,006015011 |
| 90 | 0,007647403 | 4183 | 0,005944037 |
| 3864 | 0,007611916 | 92 | 0,00585532 |

Таблица 5 – Ранжированная матрица взвешенной центральности сети

| № узла | 46 | 880 | 69 | ... | 3127 | ... | 157 | ... | 8371 |
|--------|----------|----------|---------|-----|----------|-----|----------|-----|----------|
| 46 | 0,010575 | 0 | 0 | ... | 0 | ... | 0 | ... | 0 |
| 880 | 0 | 0,009741 | 0 | ... | 0 | ... | 0 | ... | 0 |
| 69 | 0 | 0 | 0,00939 | ... | 0 | ... | 0 | ... | 0 |
| ... | ... | ... | ... | ... | ... | ... | ... | ... | ... |
| 3127 | 0 | 0 | 0 | ... | 0,001357 | ... | 0 | ... | 0 |
| ... | ... | ... | ... | ... | ... | ... | ... | ... | ... |
| 157 | 0 | 0 | 0 | ... | 0 | ... | 0,000157 | ... | 0 |
| ... | ... | ... | ... | ... | ... | ... | ... | ... | ... |
| 8371 | 0 | 0 | 0 | ... | 0 | ... | 0 | ... | 0,000018 |

Для данной сети допустима 5% потеря трафика [12], следовательно, полученная сумма больших значений степеней центральности должна быть не меньше 0,95. Этим критерием и ограничится данная репрезентативная выборка [9 – 12].

Таблица 6 – Проранжированная усеченная матрица социальной сети Rate&Goods

| Порядковый № | 1 | 2 | 3 | ... | 50 | 51 | ... | 200 | 201 | ... | 384 |
|--------------|----------|----------|---------|-----|----------|----------|-----|---------|---------|-----|----------|
| № узла | 46 | 880 | 69 | ... | 92 | 18 | ... | 157 | 9768 | ... | 47 |
| 1 | 0,010575 | 0 | 0 | ... | 0 | 0 | ... | 0 | 0 | ... | 0 |
| 2 | 0 | 0,009741 | 0 | ... | 0 | 0 | ... | 0 | 0 | ... | 0 |
| 3 | 0 | 0 | 0,00939 | ... | 0 | 0 | ... | 0 | 0 | ... | 0 |
| ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... |
| 50 | 0 | 0 | 0 | ... | 0,005856 | 0 | ... | 0 | 0 | ... | 0 |
| 51 | 0 | 0 | 0 | ... | 0 | 0,005309 | ... | 0 | 0 | ... | 0 |
| ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... |
| 200 | 0 | 0 | 0 | ... | 0 | 0 | ... | 0,00247 | 0 | ... | 0 |
| 201 | 0 | 0 | 0 | ... | 0 | 0 | ... | 0 | 0,00211 | ... | 0 |
| ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... |
| 384 | 0 | 0 | 0 | ... | 0 | 0 | ... | 0 | 0 | ... | 0,000107 |

В результате выборки получили 384 вершин (табл. 6), с помощью которых можно описать данную сеть с 5% погрешностью. Воспользуемся ПО «Gephi» [8] для визуализации усеченной выборки (рис. 10).

У данной сети выделяется три связанных кластера. По полученным данным можно утверждать, что общение между рядовыми пользователями (3) намного активнее, чем у vip-пользователей и экспертов (2), связующими звеньями между ними являются модераторы-контроллеры (1) [7], основной задачей которых является публикация предложенного пользователями контента и распространение между остальными субъектами сети.

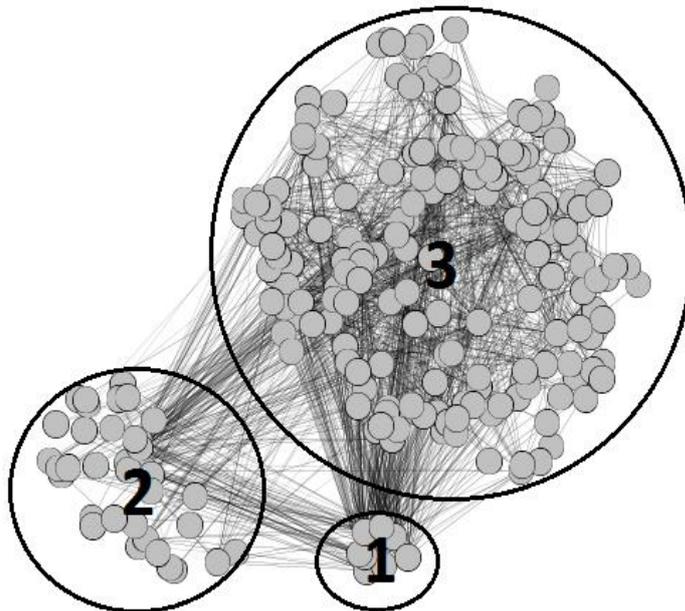


Рис. 10. Выборка социальной сети Rate&Goods

Состав усеченной выборки должен представлять (моделировать) генеральную совокупность, поскольку выводы, полученные в эксперименте, предполагается в дальнейшем перенести на всю генеральную совокупность. Поэтому выборка должна обладать особым качеством — репрезентативностью, позволяющим распространить полученные на ней выводы на всю генеральную совокупность [10].

Возможно обеспечить репрезентативность выборки на основе доказательства о принадлежности полученной выборки к степенному закону

распределения по критерию Пирсона [11]:

$$\chi^2_{\text{набл}} = \sum_{i=1}^k \frac{(n - np_i)^2}{np_i}, \quad (3)$$

где n – объем выборки;

p_i - вероятность попадания случайной величины в заданный интервал степеней.

Следуя представленной формуле (3), рассчитаем значение χ^2 . В данном случае $\chi^2=21.46$. Следовательно, $n = 21$ (выборка) – 1 (количество параметров в плотности степенного распределения) = 20 (число степеней свободы) [11].

Выберем уровень значимости $\gamma=0,05$, тогда $p=1-\gamma=0,95$. Табличное значение $\chi^2(21)$ равно 28.9. Получается, что табличное значение χ^2 больше вычисленного. Получается, что табличное значение χ^2 больше вычисленного. Следовательно, гипотеза верна, и приведенное степенное распределение согласуется с полученными результатами.

Процесс непосредственного определения репрезентативности выборки также складывается из сопоставления средних показателей распределений усеченной и генеральной совокупностей и сопоставления форм распределения этих показателей [10]. В данном случае берется среднее квадратическое отклонение:

$$\sigma = \sqrt{\frac{1}{N-1} \sum_{i=1}^N (x_i - \bar{x})^2}, \quad (4)$$

где: N – общее число вершин;

x_i – i -ый элемент выборки;

\bar{x} – среднее арифметическое выборки.

Из (4) рассчитаем среднеквадратичное отклонение (СКО) усеченной $\sigma_{УВ}$ и генеральной $\sigma_{ГС}$ выборок [11, 13]. СКО рассчитывается как квадратный корень дисперсии фактических значений переменных, определяемой по формуле (4).

Подтверждая репрезентативность, необходимо учесть, что разность полученных значений $\sigma_{ГС}$ и $\sigma_{УВ}$ должна быть по модулю меньше 5% от СКО генеральной совокупности.

$$\left. \begin{aligned} \sigma_{ГС} &= 0,0131; \\ \sigma_{УВ} &= 0,0141; \end{aligned} \right\} |\sigma_{ГС} - \sigma_{УВ}| \leq 5\%(\sigma_{ГС});$$

$$0,000635 \leq 0,00067.$$

В итоге имеем отклонение между эталоном и полученной выборкой в 4,68%. Наглядно это можно изобразить на гистограмме распределения вершин (рис. 11), оси которой являются k (степень вершины) и P(k) (вероятность попадания вершины со степенью k в заданный интервал).

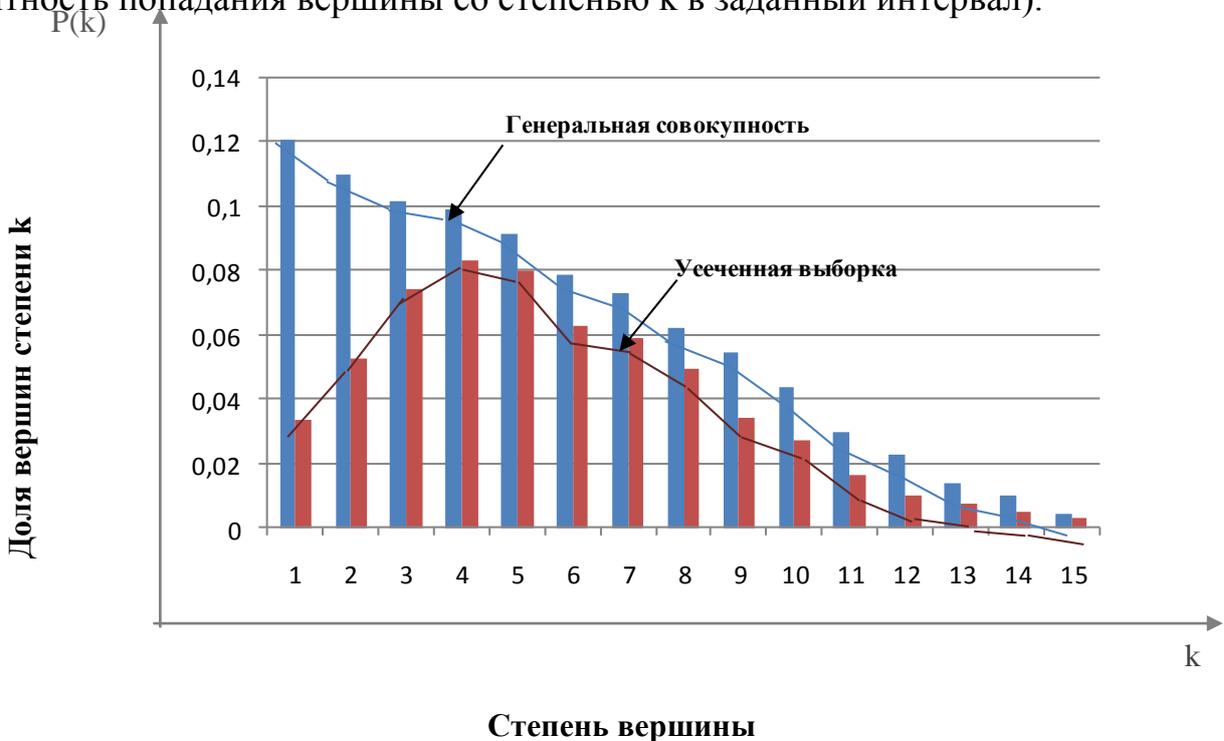


Рис. 11. Гистограмма распределения вершин

Критерий, с заданной точностью не больше 5% попадает в необходимый интервал, следовательно, полученная выборка репрезентативна [11, 12].

Вышерассмотренная выборка сети может помочь провести полный анализ исследуемого интернет-ресурса, суть которого состоит в оценке и регулировании рисков, возникающих в сети [14].

В связи с этим исследование социальных сетей и обеспечение их безопасности от вероятного преобладающего вредоносного контента на данный момент является одним из актуальных направлений как отечественных, так и зарубежных ученых [15 – 21].

Литература:

1. Анализ сайта Ratengoods.com. Дата обновления 12.06.2016 URL: <https://a.pr-cy.ru/ratengoods.com/> (Дата обращения 8.08.2016)
2. Черняк, Л. Сервисы и теории социальных сетей Текст. / Л. Черняк / Открытые системы. СУБД. – 2008. – № 8. – С. 25-31.
3. Livejournal. Russian social network. [Электронный ресурс]. – Режим доступа: <http://ru-sndm.livejournal.com/1481.html>
4. Анализ контента социальных медиа в эпоху больших данных [Электронный ресурс]. – Режим доступа: <https://mebius.io/analysis/social-media-content-analysis>
5. Posters – статистика и аналитика контента социальных сетей. Дата обновления 18.09.2016 URL: <https://www.popsters.ru/> (Дата обращения 20.09.2016).
6. Анализ социальных сетей. Дата обновления 12.06.2016 URL: <http://letopisi.org/index.php/>. (Дата обращения 1.09.2016)

7. Rate&Goods. Дата обновления 7.10.2016. URL: <http://ratengoods.com/> (Дата обращения 10.10.2016)
8. Gephi makes graphs. [Электронный ресурс]. – Режим доступа: <https://gephi.org>
9. Кто, что, где, когда: проблема выборки. [Электронный ресурс]. – Режим доступа: http://grachev62.narod.ru/Mr/Mr_05.html
10. Репрезентативность выборки. [Электронный ресурс]. – Режим доступа: http://studopedia.ru/2_31995_reprezentativnost-viborki.html
11. Гмурман В.Е. Теория вероятностей и математическая статистика. – М.: Высш. шк. – 2007. – С. 142-151.
12. Губанов Д.А., Новиков Д.А., Чхаратишвили А.Г. Социальные сети: модели информационного влияния, управления и противоборства / Под ред. чл.-корр. РАН Д.А. Новикова. – М.: Издательство физико-математической литературы. – 2010. – 288 с.
13. Боровиков В. STATISTICA. Искусство анализа данных на компьютере: Для профессионалов. [Электронный ресурс]. – Режим доступа: <http://bookre.org/reader?file=1333875&pg=17>
14. Назарчук А.В., О сетевых исследованиях в социальных науках // МГУ им. М. В. Ломоносова - М.: Типография МГУ. – 2008. – 73 с.
15. G.A. Ostapenko, D.G. Plotnicov, O.Y Makarov, N.M. Tikhomirov, V.G. Yurasov. Analytical estimation of the component viability of distribution automated information data system. World Applied Sciences Journal. – 2013. – 25 (3). – P. 416-420.
16. G.A. Ostapenko, L.V. Parinova, V.I. Belonozhkin, I.L. Bataronov, K.V. Simonov. Analytical models of information-psychological impact of social information networks on users. World Applied Sciences Journal. – 2013. – 25 (3). – P. 410-415.

17. S.A. Ermakov, A.S. Zavorykin, N.S. Kolenbet, A.G. Ostapenko, A.O. Kalashnikov. Optimization of expert methods used to analyze information security risk in modern wireless networks. Life Science Journal. – 2014. – № 11(10s). – P. 511–514.

18. N.M. Radko, A.G. Ostapenko, S.V. Mashin, O.A. Ostapenko, D.V. Gusev. Assessment of the system's EPI-resistance under conditions of information epidemic expansion. Biosciences Biotechnology Research Asia. – 2014. – Vol. 11 (3). – P. 1781–1784.

19. N.M. Radko, A.G. Ostapenko, S.V. Mashin, O.A. Ostapenko, A.S. Avdeev. Peak risk assessing the process of information epidemics expansion. Biosciences Biotechnology Research Asia. – 2014. – Vol. 11 (Spl.End). – P. 251-255.

20. A.G. Ostapenko, M.V. Bursa, G.A. Ostapenko, D.O. Butrik. Flood-attacks within the hypertext information transfer protocol: damage assessment and management. Biosciences Biotechnology Research Asia. – 2014. – Vol. 11 (Spl.End). – P. 173-176.

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

Воронежский научно-образовательный центр
управления информационными рисками

УДК 004.056

**СОЦИАЛЬНАЯ СЕТЬ ПО ИНТЕРЕСАМ LAST.FM: СТРУКТУРНО –
ФУНКЦИОНАЛЬНЫЕ ОСОБЕННОСТИ АНАЛИЗА ПРОЦЕССОВ
РАСПРОСТРАНЕНИЯ КОНТЕНТА**

Тарелкин М.А., Шварцкопф Е.А., Паринов А.В.

Социальная сеть по интересам - это сеть, которая построена по принципам связи людей, интересующихся схожей информацией в одной и той же области. Для поиска пользователей, имеющих смежные интересы, совсем не обязательно быть с ними знакомыми[1]. Ключевым здесь является проявление активности по своим интересам. Социальная сеть Last.Fm – музыкальная социальная сеть, основанная на сборе статистики прослушивания музыкальных треков[2].

Для проведения структурно-функционального анализа сети Last.Fm необходимо решить следующие задачи[3,4]:

1. Провести всестороннее исследование социальной сети Last.Fm и сформировать ее структурно-функциональную модель.

2. Провести поиск, анализ и преобразование статистических данных в соответствии с алгоритмами осуществления репрезентативной выборки моделирования диффузии контента в сети[5-7].

Анализ рассматриваемой социальной сети предусматривает выявление признаков для составления классификации, в которой ресурсы можно разделить по принадлежности коллективного или персонального пользования[8]. Коллективные ресурсы – ресурсы, в которых контент размещается или потребляется группой пользователей. Персональные ресурсы – ресурсы, в которых контент размещается каждым субъектом сети отдельно и для каждого пользователя индивидуально[9].

Профиль – персональная страница пользователя социальной сети, на которой содержится информация о ее владельце[10-12].

Форум – ресурс социальной сети для открытого обмена текстовым контентом[13].

Тема – ресурс социальной сети для открытого обмена контентом на заданную тематику[14].

Событие – ресурс, в котором публикуется информация о предстоящем мероприятии[15].

Магазин – ресурс для коммерциализации контента.

Интернет-радио – ресурс, в котором размещается наиболее популярная музыка за сутки[15].

Классификация изображена на рисунке 1.

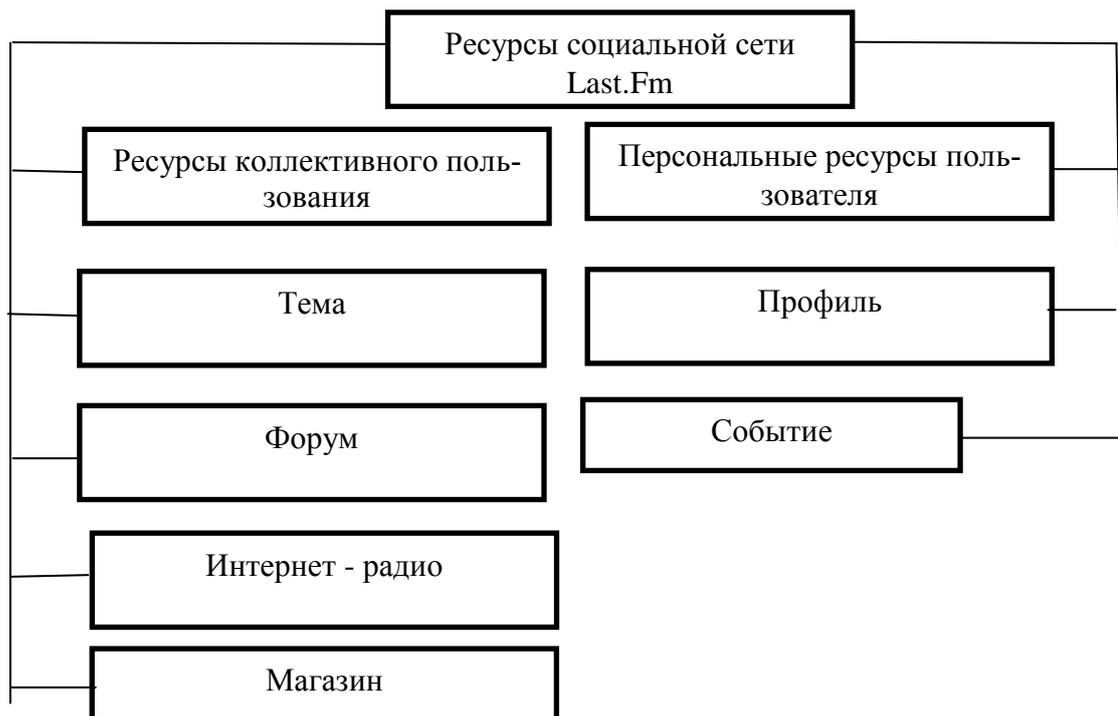


Рис. 1. Классификация ресурсов социальной сети Last.Fm

Представим классификацию контента на рисунке 2. Классификация осуществлена по признаку различной типизации файлов, являющейся специфической для рассматриваемой социальной сети.

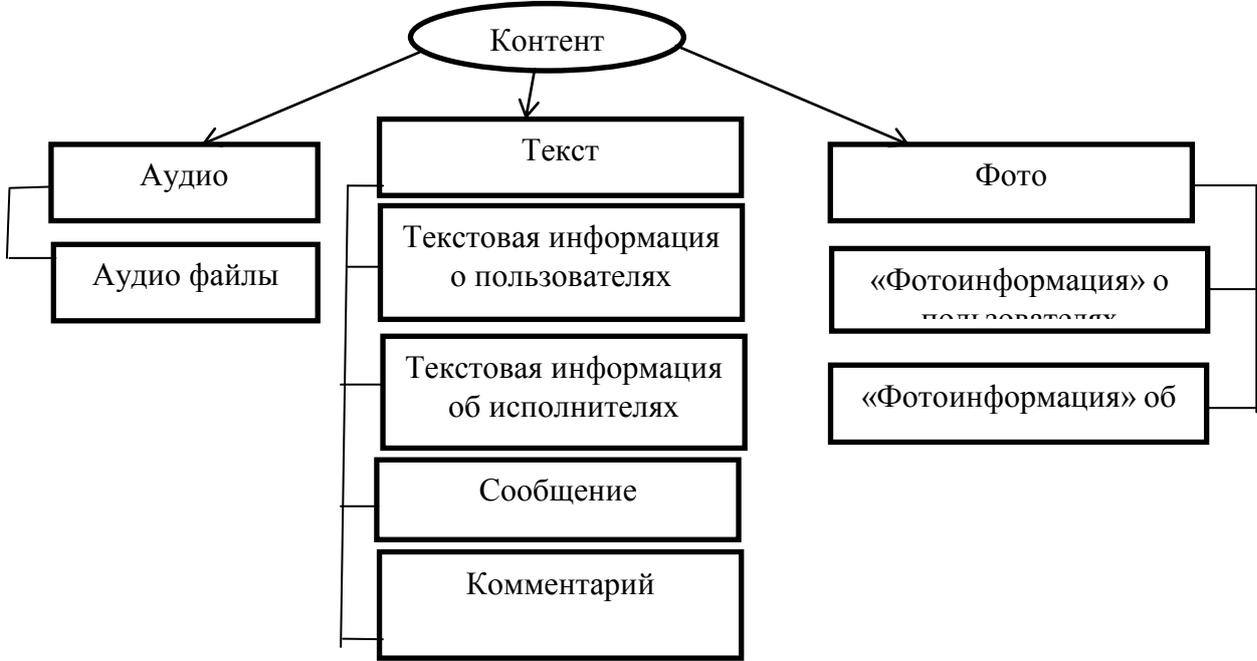


Рис. 2. Классификация контента, циркулирующего в социальной сети Last.Fm

Изобразим схематически на рисунке 3 классификацию субъектов социальной сети Last.Fm. Классификационным признаком будет являться активность или пассивность пользователей в плане размещения контента.

Авторизованный пользователь – участник социальной сети, прошедший процесс регистрации[15].

Неавторизованный пользователь – участник социальной сети, который не зарегистрирован[15].

Модератор – пользователь, имеющий более широкие права по сравнению с рядовыми участниками социальной сети[16].

Модератор вики – статей – пользователь, занимающийся созданием и редактированием статей об исполнителях с сайта Википедия.

Модератор форума – пользователь, осуществляющий контроль за исполнением правил общения на форуме.

Модератор событий – пользователь, который отслеживает концерты, презентации и другие мероприятия и предлагает их другим пользователям.

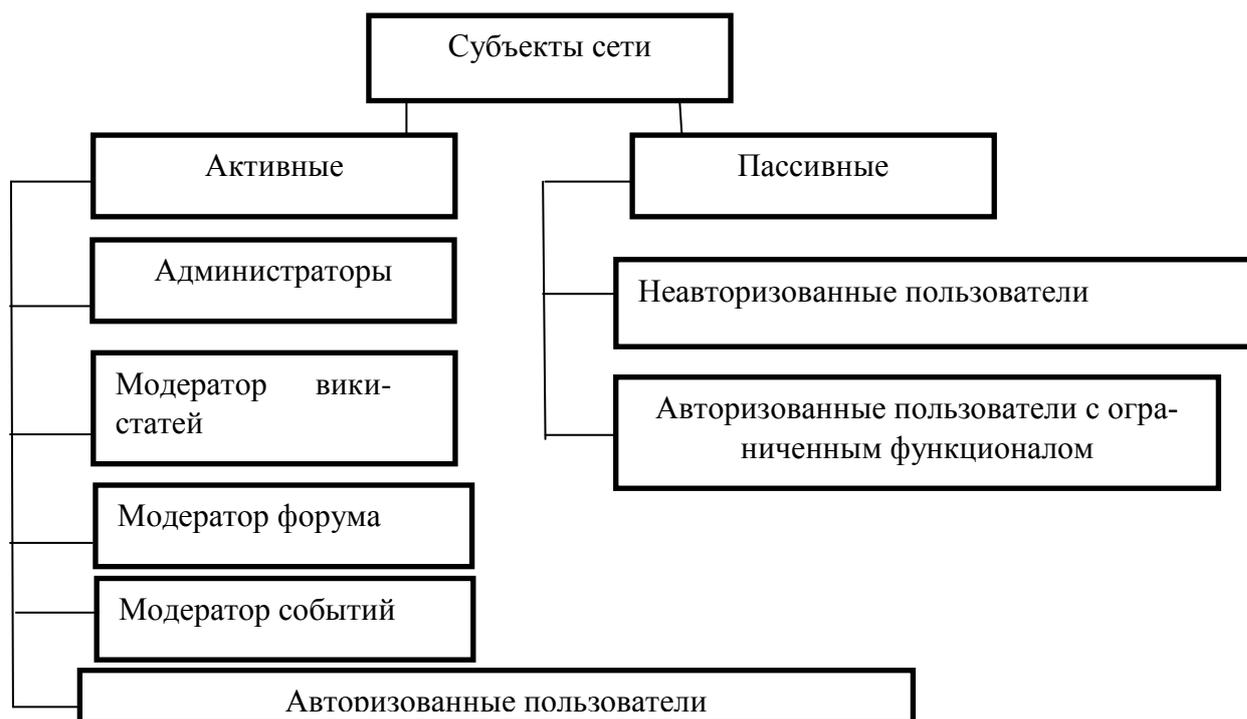


Рис. 3. Классификация субъектов, функционирующих в социальной сети Last.Fm

Администратор – участник сети, создавший сеть и управляющий ролями пользователей[17].

Автор – пользователь, который сгенерировал контент[18].

Подписчик – участник социальной сети, потребляющий контент[18].

Объектами социальной сети будем считать то, что создаётся внутри социальной сети и размещается на её ресурсах. Контент – информационно значимое наполнение ресурса социальной сети[19].

Сообщение – контент, направленный от одного пользователя другому приватно[20].

Пост – это публичная запись, опубликованная пользователем[20].

Комментарий – запись пользователя, отражающая его мнение/отношение к ранее сгенерированному контенту[21,22].

Лайк – специальный счетчик, который отражает значимость объекта для оценивающего субъекта[21].

Хэштег – это слово или фраза с префиксом#, которые отражают смысловую нагрузку помеченного контента и позволяют осуществлять его поиск[23].

Классификацию объектов социальной сети продемонстрируем на рисунке 4.

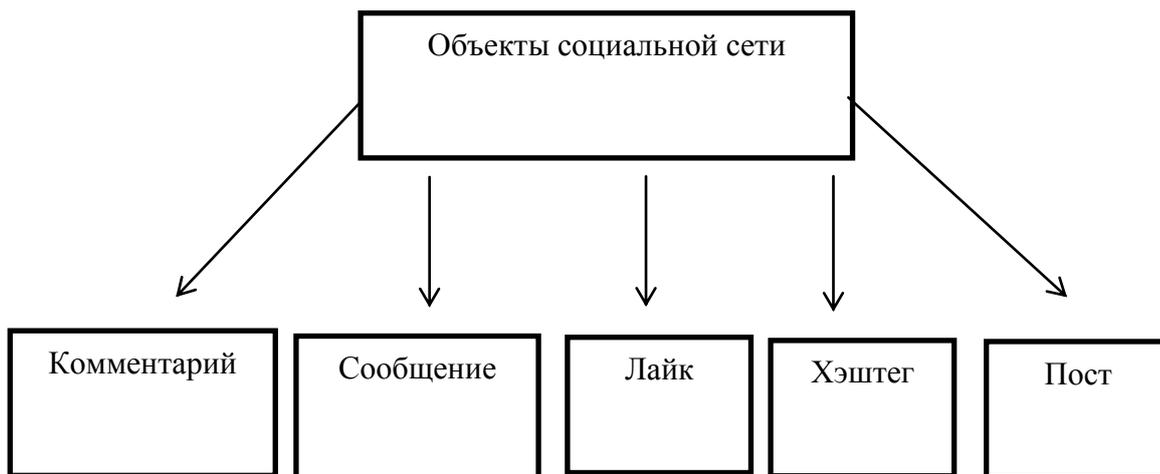


Рис. 4. Классификация объектов, создаваемых в ресурсах социальной сети Last.Fm

Действие в социальной сети – то, что может сделать субъект социальной сети с контентом, размещённым в ней. Регистрация пользователя – процедура регистрации пользователя социальной сети, в результате которой создается новая учетная запись[24].

Публикация – размещение контента на общедоступном ресурсе[24,26].

Ознакомление – просмотр ранее сгенерированного в сети контента[25,26].

Распространение – передача контента в социальной сети между пользователями.

Сохранение – передача контента в личный ресурс для его хранения.

Комментирование – создание комментария.

Оценивание – установление значимости объекта для оценивающего субъекта.

Редактирование – внесение изменений в опубликованный контент.

Подписка – однонаправленный контакт для получения обновляемого контента с определенного ресурса на регулярной основе.

Рекомендация – предложение пользователю ознакомиться с потенциально значимым для него контентом.

Обсуждение – обмен контентом между пользователями на заданную тематику.

Добавление тега - прикрепление тега к определенному объекту [27,28].

Классификацию действий покажем на рисунке 5.

Действия для большинства типов социальных сетей похожи, но стоит заметить, что не смотря на однообразие имеются ключевые особенности. Для социальной сети по интересам Last.Fm – модерирование размещаемого контента. Наличие различных модераторов (причём являющихся штатными сотрудниками сайта) для каждого вида контента не влияет на общий функционал действий, но оказывает влияние на построение структурно- функциональной схемы взаимодействия и механизмы диффузии контента.



Рис. 5. Классификация действий субъектов в социальной сети Last.Fm

С учётом полученных классификаций представляется возможным построить структурно – функциональную схему социальной сети Last.Fm. Представим её на рисунке 6.

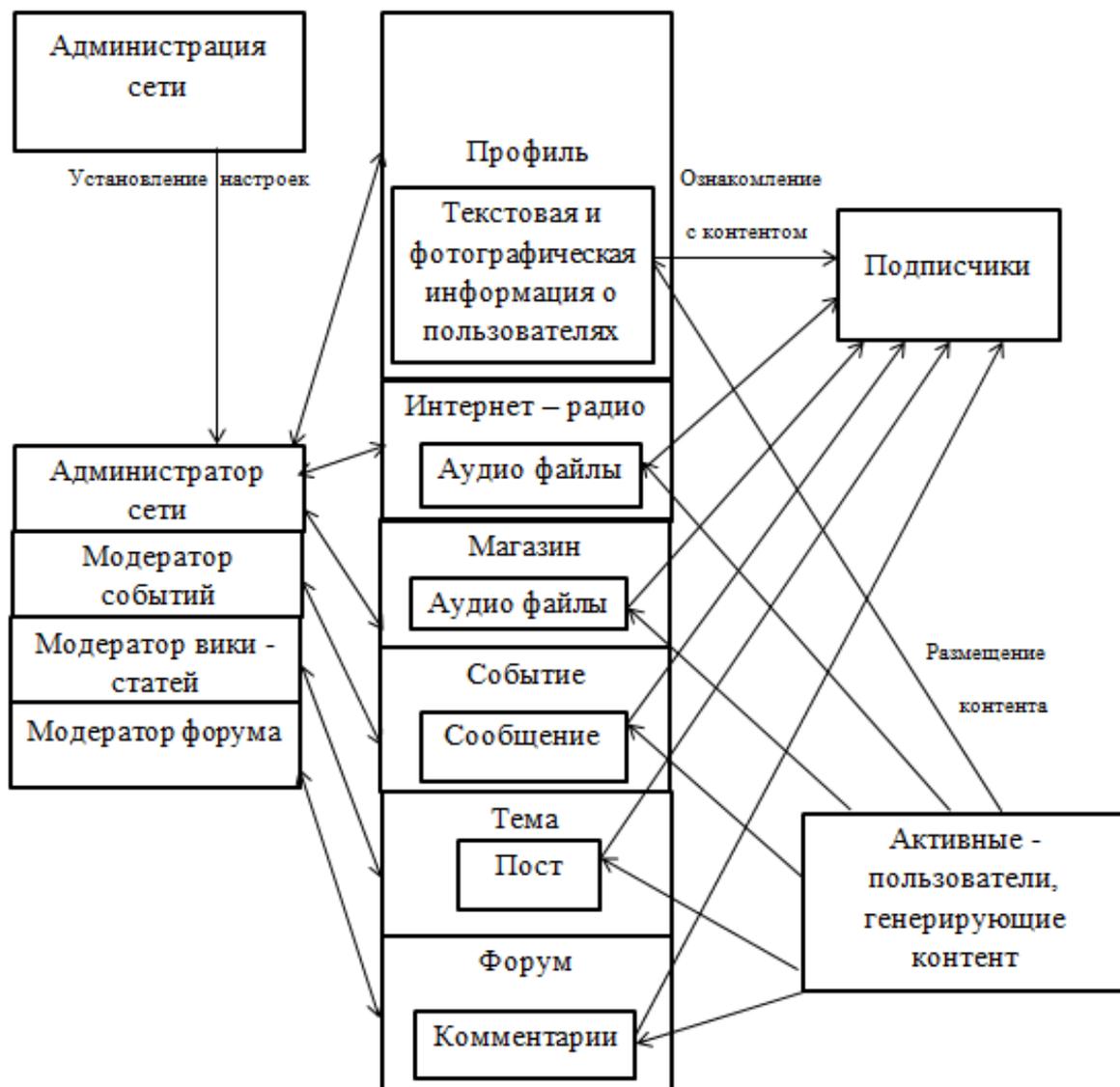


Рис. 6. Структурно – функциональная схема социальной сети Last.Fm

Преобразуем исходные статистические данные в соответствии со следующим алгоритмом. В первую очередь необходимо произвести выборку из предоставленной статистики по степени вершин, исключив низко связанные узлы социальной сети Last.Fm. Полученная сеть может быть описана при помощи метрик, изложенных в таблице 1[29]. Данные метрики будут нам необходимы для проведения структурного анализа сети.

Таблица 1 – Метрики репрезентативной выборки социальной сети Last.Fm

| | |
|--|-----------|
| Количество вершин | 58228 |
| Количество ребер | 1,150,868 |
| Ассортативность | -0.11460 |
| Диаметр сети | 6 ребер |
| Алгебраическое соединение | 0,51871 |
| Энтропия распределения относительности ребер | 73,8 % |

Далее сформируем трехместный предикат для социальной сети Last.Fm, который будет иметь вид $\Gamma(x, y, z)$, x и y - смежные вершины, z - вес инцидентного ребра для вершин x и y , z отражает количество общего контента, потребляемого двумя пользователями. В социальной сети по интересам Last.Fm вершинами являются зарегистрированные пользователи с полным функционалом (прослушали более 50 треков на сайте), а вес дуги - общее количество музыкальных треков, которые прослушали 2 пользователя (т.е. совпадение их интересов) в сутки [30]. В таблице 2 приведем фрагмент полученного трехместного предиката.

Таблица 2 – Фрагмент трехместного предиката

| Идентификатор исходящей вершины | Идентификатор входящей вершины | Вес ребра |
|---------------------------------|--------------------------------|-----------|
| 1 | 2 | 10 |
| 1 | 3 | 21 |
| 2 | 3 | 16 |
| 3 | 4 | 27 |

Представленные данные в таком предикате соответствует следующей формуле:

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

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 .

$$\langle \delta(a_{jk}) \rangle = \langle c \rangle \langle v' \rangle, \quad (2)$$

где $\langle c \rangle$ – удельная (средняя) ценность единицы объема наполнителя;
 $\langle v' \rangle$ - интенсивность обмена контентом. [31]

Значения весового коэффициента введём в таблице 3.

Таблица 3 – Весовые коэффициенты индекса восприимчивости социальной сети Last.Fm

| Социальная сеть | Практическое значение объема информации | Значение добавочного коэффициента ценности формата передаваемой информации | Используемые весовые коэффициенты индекса восприимчивости |
|-----------------|--|--|---|
| Last.fm | Количество прослушанных музыкальных треков по совпадению | - текст $k_f = 0,3$; - аудио $k_f = 0,2$; -изображение $k_f = 0,1$. | $0,1 \cdot L + 0,5 \cdot K$ |

где L – лайк;

K – комментарий.

Суть алгоритма преобразования исходных данных социальной сети Last.Fm заключается в построении тройного предиката сначала в виде звездной матрице и затем в виде матрицы взвешенной центральности. Рассматриваемый алгоритм в общем направлен на решение задачи масштабности, которая является важнейшей задачей для исследователей глобальных сетей. Дополнительную сложность добавляет условие, что свёрнутая сеть должна сохранять свойства исходной сети, отражать в наиболее полной мере её внутреннее и внешнее взаимодействие. Так же алгоритм является универсальным, применим для любого вида сетей.

На основе предложенного алгоритма преобразования исходных данных сети построим звездную матрицу социальной сети Last.Fm, элементы

строки которой соответствуют дугам, входящим в данную вершину, а элементы столбца соответствуют дугам, исходящим из вершины (таблица 4).

Таблица 4 – Звездная матрица социальной сети Last.Fm

| | | | | | | | | | | | | | | |
|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|
| | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | ... | 386 |
| 1 | - | 10 | 21 | 19 | 55 | 0 | 17 | 0 | 4 | 14 | 16 | 0 | ... | 4 |
| 2 | 0 | - | 0 | 25 | 50 | 0 | 26 | 0 | 0 | 32 | 8 | 4 | ... | 8 |
| 3 | 0 | 0 | - | 27 | 37 | 0 | 49 | 0 | 9 | 37 | 18 | 7 | ... | 0 |
| 4 | 0 | 0 | 0 | - | 43 | 0 | 52 | 0 | 14 | 45 | 14 | 0 | ... | 0 |
| 5 | 0 | 0 | 0 | 0 | - | 0 | 21 | 0 | 12 | 38 | 27 | 0 | ... | 14 |
| 6 | 0 | 0 | 0 | 0 | 0 | - | 0 | 18 | 0 | 0 | 0 | 0 | ... | 0 |
| 7 | 0 | 0 | 0 | 0 | 0 | 0 | - | 0 | 47 | 28 | 5 | 0 | ... | 0 |
| 8 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | - | 0 | 2 | 0 | 26 | ... | 0 |
| 9 | 0 | 4 | 0 | 0 | 0 | 0 | 0 | 0 | - | 8 | 4 | 0 | ... | 0 |
| 10 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | - | 20 | 3 | ... | 0 |
| 11 | 0 | 0 | 0 | 0 | 0 | 17 | 0 | 0 | 0 | 0 | - | 0 | ... | 0 |
| 12 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | - | ... | 0 |
| ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | - | 0 |
| 386 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | - |

Диагональная матрица получается обнулением всех строк и столбцов, кроме диагонали. (Таблица 5)

Таблица 5 – Диагональная матрица степеней вершин сети Last.Fm

| | | | | | | | | | | | | | | |
|-----|-----|-----|-----|-----|-----|------|-----|------|-----|-----|-----|-----|-----|-----|
| | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | ... | 386 |
| 1 | 161 | - | - | - | - | - | - | - | - | - | - | - | ... | ... |
| 2 | - | 179 | - | - | - | - | - | - | - | - | - | - | ... | - |
| 3 | - | - | 312 | - | - | - | - | - | - | - | - | - | ... | - |
| 4 | - | - | - | 225 | - | - | - | - | - | - | - | - | ... | - |
| 5 | - | - | - | - | 379 | - | - | - | - | - | - | - | ... | - |
| 6 | - | - | - | - | - | 4122 | - | - | - | - | - | - | ... | - |
| 7 | - | - | - | - | - | - | 308 | - | - | - | - | - | ... | - |
| 8 | - | - | - | - | - | - | - | 1294 | - | - | - | - | ... | - |
| 9 | - | - | - | - | - | - | - | - | 276 | - | - | - | ... | - |
| 10 | - | - | - | - | - | - | - | - | - | 303 | - | - | ... | - |
| 11 | - | - | - | - | - | - | - | - | - | - | 342 | - | ... | - |
| 12 | - | - | - | - | - | - | - | - | - | - | - | 475 | ... | - |
| ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | - |
| 386 | - | - | - | - | - | - | - | - | - | - | - | - | - | 625 |

Диагональная матрица была построена на втором шаге алгоритма, она будет необходима для последующих процедур моделирования процессов диффузии контента в социальной сети.

На третьем шаге алгоритма необходимо построить квадратную матрицу взвешенной центральности элементов сети. Для этого необходимо определить удельный вес ее вершин и дуг с помощью нормировки их весов по суммарному трафику сети по следующей формуле:

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

где $\delta(a_{ij})$ – вес трафика в дуге a_{ij} .

Данная формула будет характеризовать степень взвешенной (по трафику) центральности дуги.

Трафик сети представляет собой объём информации, передаваемой через социальную сеть за определённый период времени. Количество трафика может измеряться как в пакетах, так и в мегабайтах (МБ) и других производных.

В рамках проводимого исследования трафик социальной сети Last.Fm будем рассматривать как объём музыкального и информационного контента, переданного в единицу времени [30,31]. Необходимо учесть, что действия пользователей социальной сети Last.Fm также формируют трафик. Однако учитывая специфику социальной сети Last.Fm будем считать, что поведение пользователей неотъемлемо ведет к циркуляции контента в сети и рассматривать действия пользователей как генерацию трафика не имеет смысла.

Не оставим без внимания и тот факт, что суммарный трафик сети не разделяет входящие и исходящие дуги. Поэтому для определения взвешенной центральности вершины x_s можно использовать сумму по которой далее следует провести нормировку по суммарному трафику сети:

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

В итоге получим нормированную величину, которая будет характеризовать удельный вес трафика, проходящего через вершину x_s , по отношению ко всему трафику сети:

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

где $\delta(a_{si})$ – вес трафика в исходящей из узла a_{si} дуге;

$\delta(a_{js})$ – вес трафика во входящей в узел a_{js} дуге.

В итоге, следуя формуле (5) получим матрицу взвешенной центральности элементов исходной сети следующего вида (Таблица 6).

Таблица 6 - Матрица взвешенной центральности социальной сети Last.Fm

| | | | | | | | |
|-----|----------|-----------|-----------|-----------|----------|-----|----------|
| | 1 | 2 | 3 | 4 | 5 | ... | 386 |
| 1 | 0,001366 | - | - | - | - | - | ... |
| 2 | - | 0,0015187 | - | - | - | - | - |
| 3 | - | - | 0,0026472 | - | - | - | - |
| 4 | - | - | - | 0,0019090 | - | - | - |
| 5 | - | - | - | - | 0,003216 | - | - |
| ... | - | - | - | - | - | - | - |
| 386 | - | - | - | - | - | - | 0,005303 |

Заключительный этап алгоритма преобразования исходных данных сети заключается в нахождении и построении диагональной матрицы удельного баланса трафика в вершинах социальной сети Last.Fm.

На данном этапе будет учитываться не сумма, а разность исходящего и входящего для вершины трафиков:

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

где $\delta(a_{si})$ – вес трафика в исходящей из узла a_{si} дуге;

$\delta(a_{js})$ – вес трафика во входящей в узел a_{js} дуге.

В результате данных вычислений получим диагональную матрицу удельного баланса трафика в вершинах сети (таблица 7).

Таблица 7 - Матрица удельного баланса трафика в вершинах социальной сети Last.Fm

| | | | | | | | |
|-----|----------|----------|----------|----------|----------|-----|----------|
| | 1 | 2 | 3 | 4 | 5 | ... | 386 |
| 1 | 0,000051 | - | - | - | - | - | ... |
| 2 | - | 0,000112 | - | - | - | - | - |
| 3 | - | - | -0,00014 | - | - | - | - |
| 4 | - | - | - | 0,000001 | - | - | - |
| 5 | - | - | - | - | 0,000008 | - | - |
| ... | - | - | - | - | - | - | - |
| 386 | - | - | - | - | - | - | 0,000013 |

Таким образом, с использованием полученной матрицы удельного баланса можно сделать вывод о роли вершины социальной сети Last.Fm:

- 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.

Статистические данные взятые из базы знаний по социальным сетям предоставляют топологическую картину взаимосвязей субъектов.

Так как социальная сеть Last.Fm достаточно большая и включает в себя десятки тысяч вершин, что затрудняет в полной степени ее проанализировать, следует преобразовать всю сеть в некоторую выборку сети, которая в полной мере описывает и отражает все основные свойства сети.

Данный подход заключается в отборе вершин по их степени, то есть количеству инцидентным им дуг.

Однако такой подход не учитывает тот факт, что каждая дуга имеет свой вес.

Таким образом, отсутствует взвешенная оценка с точки зрения трафика.

Суть данного алгоритма состоит в том, что требуется из полученных квадратных матриц взвешенности, получить усеченное количество узлов которые способны с сохранением подобия описать исходную сетевую структуру.

Поэтому в предложенном алгоритме уместно использование квадратной матрицы взвешенной центральности элементов сети, а точнее – ее диагонали, где определена удельный трафик каждой вершины. Каждому из узлов выборки поставим в соответствие его удельный трафик, который является отношением всего трафика сети к трафику, проходящему через конкретный узел. Таким образом поймём значимость этого узла в масштабе всей сети. Взвешенная центральность – наиболее точный показатель используемый для оценки значимости узла сети, позволяющий без нормировки увидеть структуру сети.

Таблица 8 – Удельный трафик вершин

| Номер узла | Удельный трафик | Номер узла | Удельный трафик |
|------------|-----------------|------------|-----------------|
| 1 | 0 | 10 | 0,00173068 |
| 2 | 0,00008484 | 11 | 0,00095018 |
| 3 | 0,00017816 | 12 | 0,00033935 |
| 4 | 0,00060234 | 13 | 0,00183248 |
| 5 | 0,00156949 | 14 | 0,00163736 |
| 6 | 0 | 15 | 0,00052599 |
| 7 | 0,00139981 | 16 | 0,00073808 |
| 8 | 0,00015271 | ... | ... |
| 9 | 0,00072960 | 376 | 0,000026295 |

В силу того, что для дальнейшего анализа необходимо усечь выборку, проведем ранжирование степени узлов матрицы по убыванию и отсечем малозначимые узлы. с учетом 5% потери трафика. Проведем суммирование степеней узлов с минимальными значениями, полученная сумма должна быть не больше 0,05.

$$\delta(\bar{x}_s) = 0,000026295 + 0,000037145 + 0,000039041 + \dots \\ + 0,008645 + 0,008945 = 0,0491.$$

В результате проведённого усечения получим результаты, приведенные в таблице 9. Сравнение двух выборок с 5% и 95% процентной суммой взвешенной центральности показывает, что две выборки, составляющие генеральную совокупность подобны, значит выборка взятая по степенному закону является репрезентативной, что доказано на рисунке 7.

Под репрезентативностью в данном контексте подразумевается подобие выборки основному множеству по необходимым для исследования критериям.

Таблица 9 – Ранжирование вершин по удельному трафику вершин

| Номер узла ранжированной выборки | Номер узла усеченной выборки | Удельный трафик | Номер узла ранжированной выборки | Номер узла усеченной выборки | Удельный трафик |
|----------------------------------|------------------------------|-----------------|----------------------------------|------------------------------|-----------------|
| 1 | 6 | 0,03497369 | 10 | 97 | 0,00751739 |
| 2 | 73 | 0,01481418 | 11 | 121 | 0,00729679 |
| 3 | 71 | 0,01292211 | 12 | 81 | 0,00727982 |
| 4 | 8 | 0,01097912 | 13 | 109 | 0,0070168 |
| 5 | 79 | 0,01019005 | 14 | 181 | 0,00697437 |
| 6 | 174 | 0,01011369 | 15 | 23 | 0,00683013 |
| 7 | 68 | 0,01008824 | 16 | 17 | 0,00681316 |
| 8 | 179 | 0,0093925 | ... | ... | ... |
| 9 | 70 | 0,00783132 | 311 | 335 | 0,00094179 |

Проведем проверку полученных результатов обратным счетом, полученная сумма должна быть не меньше 0,95. Этим критерием и ограничится репрезентативная (с точки зрения трафика) выборка:

$$\delta(\bar{x}_s) = 0,0788666 + 0,0608231 + 0,046594 + 0,0437465 \\ + 0,0436489 + \dots + 0,0000118 = 0,9508.$$

Далее обнулим столбцы и строки, на пересечении которых стояла исключенная вершина и построим матрицу, отражающую связность компонентов выборки в таблице 10.

Таблица 10 – Усеченная матрица социальной сети Last.Fm

| | | | | | | | | | | |
|-----|-----|-----|-----|-----|-----|-----|-----|------|-----|-----|
| | 6 | 8 | 68 | 71 | 73 | 79 | 174 | 179 | ... | 335 |
| 6 | - | 18 | 51 | 35 | 28 | 19 | 0 | 1061 | ... | 45 |
| 8 | 0 | - | 0 | 38 | 51 | 0 | 5 | 0 | ... | 11 |
| 68 | 0 | 0 | - | 12 | 16 | 35 | 0 | 0 | ... | 10 |
| 71 | 0 | 0 | 0 | - | 67 | 0 | 0 | 4035 | ... | 0 |
| 73 | 0 | 0 | 0 | 0 | - | 0 | 66 | 0 | ... | 0 |
| 79 | 0 | 0 | 0 | 0 | 0 | - | 0 | 0 | ... | 0 |
| 174 | 0 | 0 | 0 | 0 | 0 | 0 | - | 0 | ... | 0 |
| 179 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | - | ... | 14 |
| ... | ... | ... | ... | ... | ... | ... | ... | ... | - | 12 |
| 335 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | - |

Проведем сравнение полученной усеченной выборки с начальной выборкой (рисунок 7) графическим способом. Для этого сгруппируем веса двух сортированных по убыванию выборок, проведем суммирование весов (таблица 11) и представим полученные результаты на рисунке 8.

Рассчитаем среднеквадратическое отклонение для элементов генеральной совокупности и выборки.

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

$\sigma_1=0.372$ – СКО генеральной совокупности;

$\sigma_2=0.357$ – СКО выборки;

$$\alpha = \frac{\sigma_1 - \sigma_2}{\sigma_1} \quad (8)$$

В результате вычислений получили значение $\alpha=0,0413$. Из этого можно сделать вывод, что выборки подобны с менее чем 5 % отклонением. Покажем это соотношение на графике (рисунок 9). Данная величина является допустимой ошибкой выборки для расчётов, осуществляемых в данной работе.

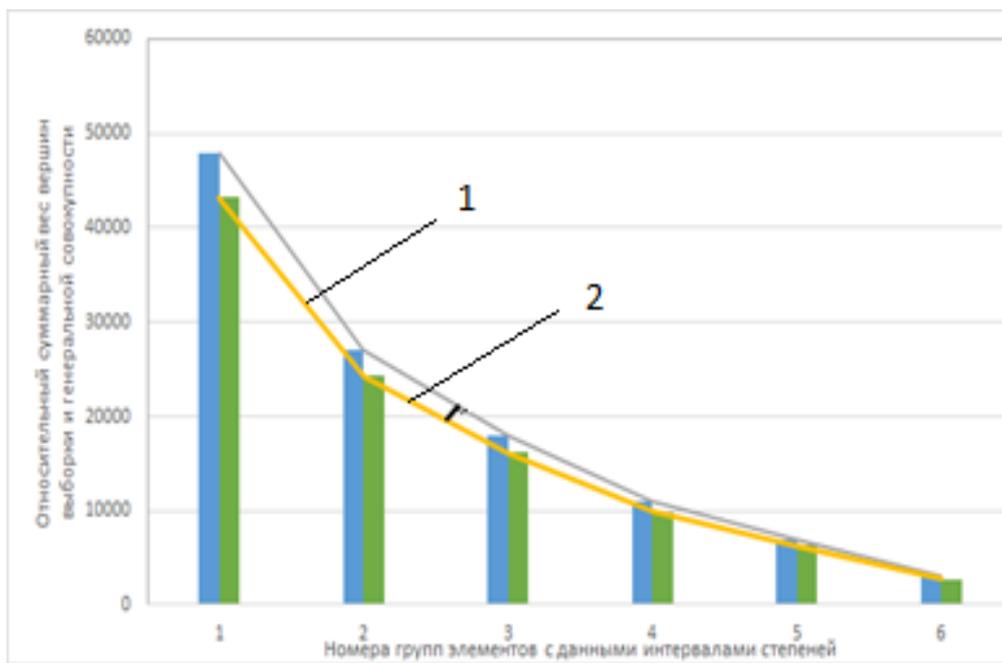


Рис. 7. Соотношение распределения генеральной совокупности и выборки.

На рисунке 7, 1 – огибающая распределения степени вершин генеральной совокупности, 2 – огибающая распределения степени вершин репрезентативной выборки.

Полученные результаты подтверждают подобие усеченной выборки исходным данным социальной сети Last.Fm.

На рисунке 8 изобразим исходную топологию социальной сети Last.Fm.

Топология говорит нам о том, что в сети имеются два крупных кластера, которые обозначают группы пользователей, потребляющих один и тот же контент, выделенный во множестве определенным образом.

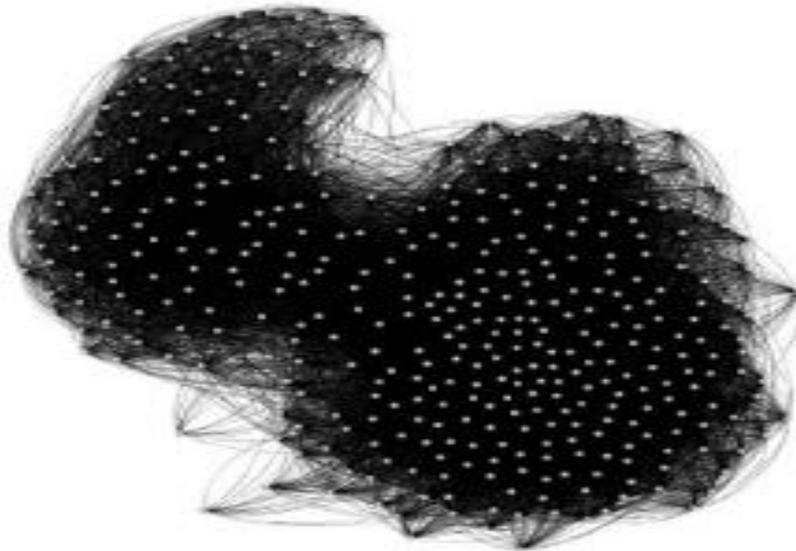


Рис. 8. Схематичное изображение исходного множества социальной сети Last.Fm, полученное при помощи ПО Gephi.

На основе полученной статистики построим усеченный граф исследуемого участка социальной сети Last.Fm при помощи программного обеспечения Gephi (<https://gephi.org/>). Полученные результаты представлены рисунком 9.

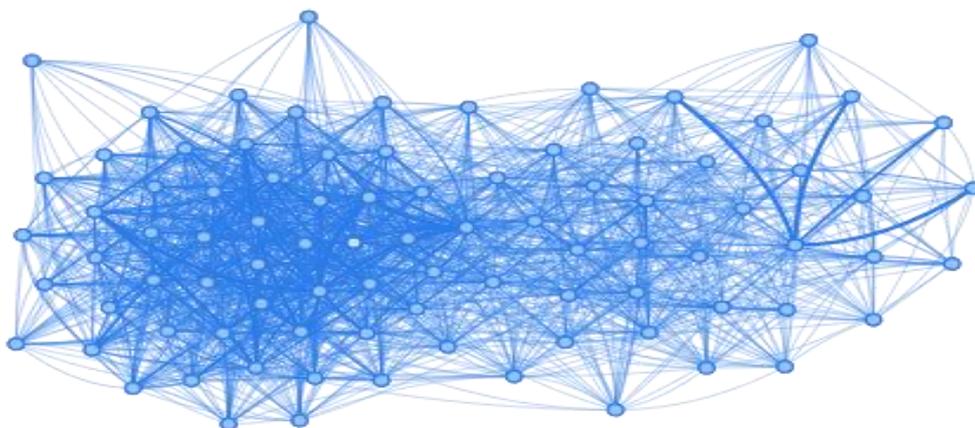


Рис. 9. Схематичное изображение выборки социальной сети Last.Fm, полученное при помощи ПО Gephi

Топология графа показывает предпочтения пользователей в плане предпочтения музыкальных направлений – тэгов. Тэг «рок» используют приблизительно 75% пользователей, а тэг поп - 25% пользователей.

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

- недиагональные элементы равны ± 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 11 – Модифицированная матрица смежности социальной сети Last.Fm

| | | | | | | | | | | |
|-----|------|------|------|------|------|------|------|------|-----|-----|
| | 6 | 8 | 68 | 71 | 73 | 79 | 174 | 179 | ... | 335 |
| 6 | 4122 | 1 | 1 | 1 | 1 | 1 | 0 | 1 | ... | 1 |
| 8 | 0 | 1294 | 0 | 1 | 1 | 0 | 1 | 0 | ... | 1 |
| 68 | 0 | 0 | 1189 | 1 | 1 | 1 | 0 | 0 | ... | 1 |
| 71 | 0 | 0 | 0 | 1523 | 1 | 0 | 0 | 1 | ... | 0 |
| 73 | 0 | 0 | 0 | 0 | 1746 | 0 | 1 | 0 | ... | 0 |
| 79 | 0 | 0 | 0 | 0 | 0 | 1201 | 0 | 0 | ... | 0 |
| 174 | 0 | 0 | 0 | 0 | 0 | 0 | 1192 | 0 | ... | 0 |
| 179 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1107 | ... | 1 |
| ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | 1 |
| 335 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 111 |

В результате применения предложенных алгоритмов преобразования исходных данных сети и нахождения репрезентативной выборки была получена визуальная модель усеченной социальной сети Last.Fm и соответствующие матрицы, способные в полной мере охарактеризовать распространение контента в данной социальной сети.

Литература:

- 1 Назарчук А.В., О сетевых исследованиях в социальных науках / МГУ им. М. В. Ломоносова - М.: Типография МГУ, 2008. – 73 с.
- 2 Anthonisse J. M., The rush in a directed graph / J.M. Anthonisse // Technical Report BN 9/71. – 1971 p. Ahn Y. Analysis of topological characteristics of huge online social networking services / Y. Ahn, S. Han, H. Knak, S. Moon, H. Jeong // 16th International Conference on the World Wide Web. – 2007. – P. 835-844
- 3 Alba R.A., graph-theoretic definition of a sociometric clique / Richard D. Alba / Journal of Mathematical Sociology. – 1973. – P. 113–126.
- 4 Губанов, Д.А. Модели информационного влияния и информационного управления в социальных сетях / Д.А Губанов, Д.А. Новиков, А.Г. Чхартишвили // Проблемы управления. – 2009. – №5. – С. 28-35.
- 5 Jennifer Golbeck. Introduction to Social Media Investigation: A Hands-on Approach. Waltham: Elsevier Inc., 2015.
- 6 Alan E. Mislove. Online Social Networks: Measurement, Analysis, and Applications to Distributed Information Systems. Houston, Texas: RICE University, 2009.
- 7 Абрамов К. Г., Моделирование распространения нежелательной информации в социальных медиа / К.Г. Абрамов, Ю.М. Монахов; Труды XXX Всероссийской научно-технической конференции. Проблемы эффективности и безопасности функционирования сложных технических и информационных систем / Серпуховский ВИ РВ. - 2011. – ч. IV. - С. 178-182.
- 8 Абрамов К. Г., Распространение нежелательной информации в социальных сетях Интернета / К.Г. Абрамов, Ю.М. – С.45-48.

- 9 Тищенко В. И. Социальные сети и виртуальные сетевые сообщества
- 10 Статистика социальной сети Last.Fm. - Электрон. Дан. – Режим доступа: <http://konect.uni-koblenz.de/networks/last.fm> (дата обращения: 07.09.2016)
- 11 IEEE Trans. on Automatic Control. – 2005. – Vol. 50, N 5. – P. 655–661.
- 12 Johnson S. Entropic origin disassortativity in complex networks / S. Johnson, J.J. Torres, J. Marro, M.A. Muñoz /Physical Review Letters. – 2010. – 4 p.
- 13 Kauai HI: IEEE. Freeman, L. C. (1979). Centrality in social networks conceptual clarification. Social Networks, 1(3), P. 215–239.
- 14 Социальная сеть . - Электрон. Дан. – Режим доступа: <http://last.fm.com> (дата обращения: 06.08.2016)
- 15 Средство визуализации данных. - Электрон. Дан. – Режим доступа: <https://gerphi.org/> (дата обращения: 15.09.2016)
- 16 Antsupov, A. Ya. Conflictology: the textbook for higher education institutions / A.Ya. Antsupov, A. I. Shilov. – 3rd prod., reslave. and additional – SPb.: St. Petersburg, 2007. – 591 p.
- 17 Bachilo, I. L. About the concept of legal support of informatization of Russia. Legislative problems of informatization of society/I.L. Bachilo, G. V. Belov / – M.: World, 1992. – 400 p.
- 18 Bachilo, I. L. Information right. The textbook Under the editorship of the academician of RAS B. N. Topornina/I.L. Bachilo, V. N. Lopatin, M. A. Fedotov and other / SPb.: Legal center, 2001. – 789 p.
- 19 Bachilo, I. L. Information right: fundamentals of practical informatics: Education guidance / I.L.Bachilo–M: Edition of . M.Yu.Tikhomirov, 2001 – 352 p.

- 20 Pálovics R. Temporal influence over the Last.fm social network Institute for Computer Science and Control / R. Pálovics // – Hungarian Academy of Sciences (MTA SZTAKI) – 2011. – 8 p.
- 21 Zhong C. Social Bootstrapping: How Pinterest and Last.fm Social Communities Benefit by Borrowing Links from Facebook / C. Zhong, M. Salehi, S. Shah // – 2014. – 14 p.
- 22 Voldstad R. Community Detection on Last.fm Artist Data / R. Voldstad // – 2014. – 9 p.
- 23 Konstas I. On Social Networks and Collaborative Recommendation / I. Konstas // – 2012. – 5p.
- 24 Maxwell A. Pretzlav Last.fm Explorer: An Interactive Visualization of Hierarchical Time-Series Data / A. Maxwell // – 2008. – 11p.
- 25 Byrd, K. War with many unknowns / K. Byrd//Computerra. – M.: 2009. – No. 20. – 5 p.
- 26 Grinyaev, S. Russia in global information society: threats, risks and possible ways of their neutralization / S. Grinyaev, – Electron. it is given. – Access mode: http://www.noravank.am/upload/pdf/419_ru.pdf.
- 27 Kalashnikov, A.O. Attacks to information and technological infrastructure of crucial objects: assessment and regulation of risks: Monograph / A.O. Kalashnikov, E. V. Yermilov, O. N. Choporov, K. A. Razinkin, N. I. Barannikov; under the editorship of the Member correspondent of RAS D. A. Novikov. - Voronezh: Scientific Book publishing house. 2013. - 160 p.
- 28 Majumdar A. Music Recommendations based on Implicit Feedback and Social Circles: The Last FM Data Set / A. Majumdar, A. Kumar, S. Manohar// – 2009. – 12 p.
- 29 Aiello L.M. Friendship prediction and homophily in social media / L.M. Aiello // – 2008. – 40 p.

30 Carneiro M. Towards the discovery of temporal patterns in music listening using Last.fm profiles / M. Carneiro // – 2011. – 71 p.

31 Гмурман В.Е., Теория вероятностей и математическая статистика. Учебное пособие. Высшее образование. – Москва, 2006 – С.243

32 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.

33 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.

34 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.

Воронежский научно-образовательный центр
управления информационными рискам

**ШАНСЫ И РИСКИ АТАКУЕМЫХ СЕТЕВЫХ ЭЛЕМЕНТОВ ДЛЯ
АНАЛИТИЧЕСКОЙ ОЦЕНКИ ХАРАКТЕРИСТИК ИХ
ЖИВУЧЕСТИ**

Д.Г. Плотников, В.В. Морковина, А.А. Остапенко

Живучесть сетей складывается из параметрической живучести элементов, прежде всего вершин их графа, т. е. — пользователей в сетевой структуре, которые подвергаются всевозможным атакам извне. Шансы успешной работы и риски отказов, возникающих в данном случае, требуют своей аналитической оценки. В настоящей работе предпринята такая попытка с учетом трафика, циркулирующего в единицу времени через вершину сети наполнителя, т. е. динамический ресурс смежных с вершиной дуг.

Первоначально уместно ввести некоторые обозначения и определения:

1. Динамический ресурс вершины сети $Res'(x_i)$ представляет собой сумму динамических ресурсов дуг, инцидентных вершине x_i

$$Res'(x_i) = \sum_j Res'(a_{ij}) + \sum_k Res'(a_{ki}).$$

Фактически этот параметр оценивает объем наполнителя сети, прокачиваемого в единицу времени через вершину x_i . Он позволяет установить связь вершины с сетевым графиком через нормирование

$$Res'(x_i) = Res'(x_i) / \sum_k Res'(a_{jk}).$$

по общему динамическому ресурсу сети (сумме динамических ресурсов всех входящих в сеть дуг). Исходными данными подобного расчета

могут служить для социальных сетей Интернет-сведения об их коммуникациях. Для корпоративных сетей следует использовать статданные о внутрисетевом графике.

2. Штатное значение динамического ресурса вершины сети $Res'_{шт}(x_i)$ относится к штатному периоду работы элемента (t_i, t_i) , исключая извне жизненного цикла (рис.1,а) переходы запуска $(0, t_i)$ и старения $(t_2, T_{ср})$, где $T_{ср}$ — средняя продолжительность жизни элементов сети подобного типа. В отсутствие внешнего воздействия (атаки на элемент) график временной зависимости динамического ресурса вершины сети будет иметь вид, представленный на рис.1,б. В случае успеха атаки на элемент в момент t_a возникает потеря ресурса. Он снижается до нештатного значения $Res'_H(x_i)$, т. е. происходит частичная потеря работоспособности (рис. 1,в). В случае полной потери работоспособности график принимает вид, проиллюстрированный на рис.1,г. В обоих случаях за период наблюдения (t_1, t_2) восстановление работоспособности элемента не происходит. Эта атака приводит к отказу без восстановления.

3. В случае, когда в момент t_b происходит восстановление динамического ресурса элемента сети x_i , график его временной зависимости примет вид (рис. 2,а). Полный отказ в период (t_a, t_b) иллюстрирует рис.2,б. В период наблюдения возможно несколько таких атак (рис. 2,в), в том числе с частичной утратой работоспособности (рис. 2,г) в периоды (t_{a1}, t_{b1}) и (t_{a2}, t_{b2}) .

Вышеизложенные соображения могут послужить основой для оценки ущерба. Для приведенных выше графиков имеем:

1) Из рис. 1,в ущерб будет равен:

$$u(x_i) = [Res'_{шт}(x_i) + Res'_H(x_i)](t_2 - t_a).$$

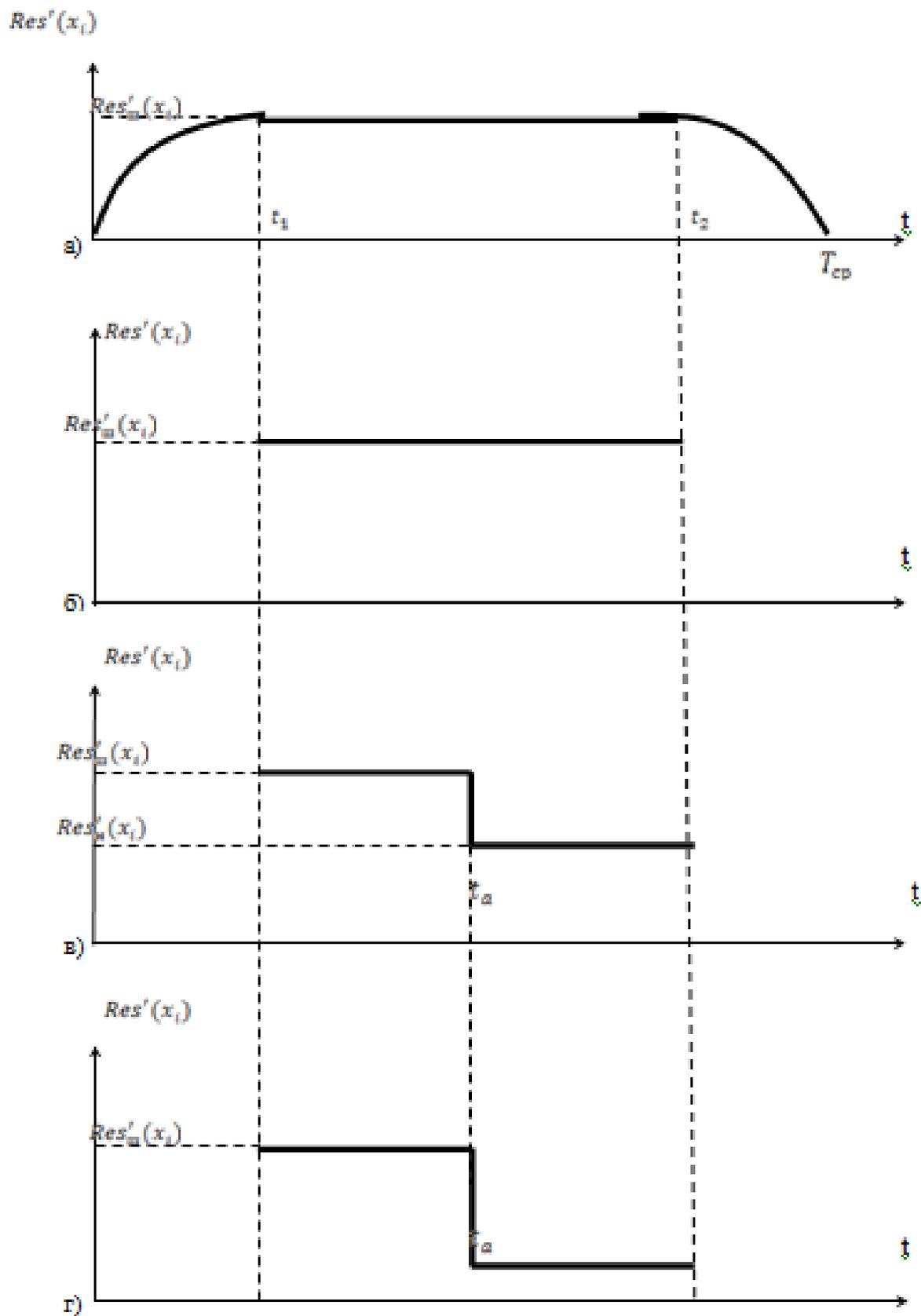


Рис. 1. Графики временной зависимости динамического ресурса вершины сети при атаке без восстановления

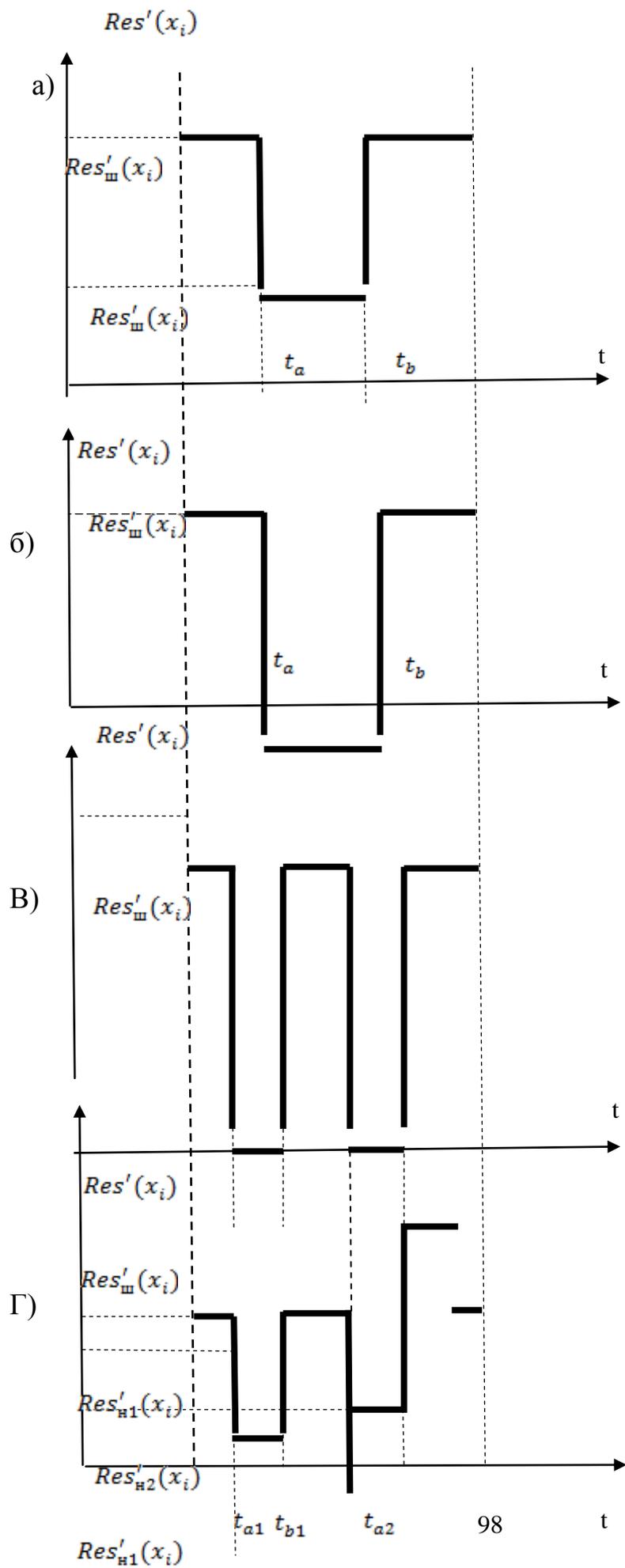


Рис. 2 — Графики временной зависимости динамического ресурса вершины сети при атаке с восстановлением

- 2) Ущерб для графика (рис.1, 2) составит:

$$u(x_i) = Res'_i(x_i)(t_2 - t_a).$$

- 3) По рисунку 2,а ущерб может быть найден следующим образом:

$$u(x_i) = [Res'_ш(x_i) - Res'_н(x_i)](t_b - a).$$

- 4) Для рисунка 2,б по аналогии получим:

$$u(x_i) = Res'_ш(x_i)(t_b - t_a).$$

- 5) Из рисунка 2,в следует, что ущерб будет равен:

$$u(x_i) = Res'_ш(x_i)[(t_{b1} - t_{a1}) + (t_{b2} - t_{a2})].$$

- 6) По аналогии, для рис. 2, г имеем:

$$u(x_i) = [Res'_ш(x_i) - Res'_н(x_i)](t_{b1} - t_{a1}) + [Res'_ш(x_i) - Res'_н(x_i)](t_{b2} - t_{a2}).$$

Фактически вышеприведенный анализ сводится к вычислению площадей, утраченных в графиках (рис. 1,в,г и рис. 2) в результате успешных атак. Очевидно этот расчет имеет приближенный характер, ибо не учитываются переходные процессы отказов и восстановления работоспособности элемента. Однако, если исходить из дискретизации времени в наблюдаемый период (t_1, t_2) шагом, значительно большим постоянной времени вышеупомянутых переходных процессов, такое приближение представляется вполне допустимым для практики.

Уместно отметить, что подобный подход можно использовать и для оценки пользы, полученной от элемента в ходе его эксплуатации. В этом случае:

- 1) Для рисунка 1,в имеем:

$$v(x_i) = Res'_ш(x_i)(t_a - t_1) + Res'_н(t_2 - t_a).$$

- 2) По аналогии для рисунка 1,г получаем:

$$v(x_i) = Res'_ш(x_i)(t_a - t_1).$$

- 3) Польза для элемента с графиком (рис. 2,а) будет равна:

$$v(x_i) = Res'_ш(x_i)[(t_a - t_1) + (t_2 - t_b)] + Res'_н(t_b - t_a).$$

4) По аналогии для рисунка 2,б польза будет равна:

$$v(x_i) = Res'_{\text{ш}}(x_i)[(t_a - t_1) + (t_2 - t_b)].$$

5) Для множества атак (рис. 2, в) имеем:

$$v(x_i) = Res'_{\text{ш}}(x_i)[(t_{a1} - t_1) + (t_{a2} - t_{b1}) + (t_2 - t_{b2})].$$

6) Соответственно при неполном отказе для множества атак польза эксплуатации элемента в период его жизненного цикла составит:

$$v(x_i) = Res'_{\text{ш}}(x_i)(t_{a1} - t_1) + Res'_{\text{ш1}}(x_i)(t_{b1} - t_{a1}) + Res'_{\text{ш}}(x_i)(t_{a2} - t_{b1}) + Res'_{\text{ш2}}(x_i)(t_{b2} - t_{a2}) + Res'_{\text{ш}}(x_i)(t_2 - t_{b2})$$

Приведенные выше выражения фактически измеряют площадь под кривой. Так оценивается польза.

После того, как измерения пользы и ущерба осуществлены, представляется возможным найти аналитические выражения для шансов штатной работы и рисков возникновения отказов элемента x_i . Для этого надо воспользоваться графиками вероятностей успеха атаки на этот элемент (рис. 3). В частности, для шанса можно записать

$$Chance(t_a) = [1 - F(t_a)]v(t_1, t_a), (1)$$

где именно для x_i фактически находится вероятность его штатной работы $[1 - F(t_a)]$ до момента атаки t_a , $av(t_1, t_a)$ – есть польза полученная за этот период.

В свою очередь риск может быть описан следующим образом [11-13]

$$Risk(t_a) = F'(t_a)u(t_a, t_2)(\Delta t), (2)$$

где $F'(t_a)$ – плотность вероятности отказа в момент атаки t_a и Δt – шаг дискретизации в период наблюдения (t_1, t_2) , а $u(t_a, t_2)$ есть ущерб, который выражается в утрате трафика после атаки.

Выражения (1) и (2) относятся к графику (рис.1,в), но и для других иллюстраций (рис. 1,г; рис. 2) подобные выражения уместно и возможно получить.

Из этих выражений можно найти ценную серию характеристик живучести, в том числе — отношение шанс/риск (прогнозируемая эффективность защиты узла)

$$E_f(x_i, t_a) = \frac{Chance(t_a)}{Risk(t_a)} \quad (3)$$

в точке t_a , а также во всем диапазоне наблюдения,

$$E_{f\Sigma}(x_i) = \frac{\sum_{t_a=t_1}^{t_2} Chance(t_a)}{\sum_{t_a=t_1}^{t_2} Risk(t_a)}$$

т. е. расчет (3) для множества временных дискрет.

Для множества синхронно атакуемых элементов сети очевидно потребуется разработка методики, учитывающей степень взвешенной центральности каждой вершины сети, в которой атака оказалась успешной. Здесь представляется возможность перехода к теории структурной живучести

В этом случае нормирование динамического ресурса узла следует осуществлять по таковому показателю для всей сети, т.е.

$$\overline{Res'}(x) \quad (4)$$

где $Res'(X) = \Sigma Res'(x_i)$.

Отсюда получается и нормированный риск[11-13]

$$\overline{Risk}_\Sigma(X_a, t_a) = \sum_i Res'_\text{ш}(x_i)(t_b - t_a)P_i(t_a)h_i(t_a), \quad (5)$$

где $P_i(t_a)$ - частота статистически ожидаемых отказов элемента сети x_i в момент его жизненного цикла t_a ;

$h_i(t_a) = Res'_\text{ш}(x_i, t_a) / Res'_\text{н}(x_i, t_a)$ – глубина утраты трафика при атаке;

\sum_i - суммирование по всему множеству успешно атакованных элементов сети $x_i \in X_a$;

$Res'_{\text{ш}}(x_i)$ - определяется выражением (4) и фактически является степенью центральности элемента в сети с точки зрения ее трафика.

Разумеется, возможно подобную мультиэлементную оценку предложить и для шанса, где

$$\overline{Chance}_{\Sigma}(X_a, t_a) = 1 - \overline{Risk}_{\Sigma}(X_a, t_a). \quad (6)$$

Выражение (6) с учетом (3) дает прогноз эффективности как отношение

$$E_f(X_a, t_a) = \frac{\overline{Chance}_{\Sigma}(X_a, t_a)}{\overline{Risk}_{\Sigma}(X_a, t_a)}$$

и в точке t_a или на всем наблюдаемом интервале

$$E_{f\Sigma}(X_a) = \frac{\sum_{t_a=t_1}^{t_2} \overline{Chance}_{\Sigma}(X_a, t_a)}{\sum_{t_a=t_1}^{t_2} \overline{Risk}_{\Sigma}(X_a, t_a)}. \quad (7)$$

Выражение (7) интересно для предсказания эффективности защиты (защищенности) элемента сети от имеющихся сетевых атак [14]. Его практическое применение возможно при наличии соответствующей P_i статистики частот отказа, т.е. элементной живучести сети. Таковая, например, имеется в реально функционирующих беспроводных сетях, где оценка их ожидаемой защищенности имеет важное значение для успешного функционирования этих сетевых структур и прочих информационных систем [1-8,10].

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. Radko, R.K. Babadzhanov, O.A. Ostapenko // Journal of Theoretical and Applied Information Technology. – 2016. – P. 306-315.

8. Дешина А.Е. Атакуемый web-сервер: оценка ожидаемой живучести [Текст] / А.Е. Дешина, Д.Г. Плотников, Е.С. Бурдастых, П.А. Анцупов, Е.А. Москалева // 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. 9. – № 3. – С. 70-80.

9. Дешина А.Е. Информационные риски в мультисерверных системах: атаки комплексного типа [Текст] / А.Е. Дешина, В.И. Белоножкин // Информация и безопасность. – 2013. – Т. 16. – № 3. – С. 335-344.

10. Остапенко Г.А. Алгоритмизация оценки живучести сетевых информационных структур [Текст] / Г.А. Остапенко, Я.С. Мишина, В.И. Белоножкин, И.В. Шевченко // Информация и безопасность. – 2014. – Т. 17. – № 2. – С. 304-307.

11. Паринов А.В. Риск-оценка смертности инновационных проектов: научно-методические основы [Текст]/ А.В. Паринов, Л.В. Паринова, В.Г. Юрасов // Информация и безопасность. – 2013. – Т. 16. – № 3. – С. 423-426.

12. Паринов А.В. Управление рисками инновационных проектов [Текст]/ А.В. Паринов, Л.В. Паринова, А.В. Заряев // Информация и безопасность. – 2014. – Т. 17. – № 4. – С. 602-605.

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. 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

Научное издание

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

Сборник научных трудов

Вып. 2 (12), 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.