

УДК 519.1 + 519.6

ББК 22.176

## ЗАДАЧА ДЕЛЕЖА ЗАТРАТ НА СОЗДАНИЕ ВЕБ-КОММУНИКАТОРА КАК КООПЕРАТИВНАЯ ИГРА<sup>1</sup>

Мазалов В. В.<sup>2</sup>, Печников А. А.<sup>3</sup>, Чирков А. В.<sup>4</sup>, Чуйко Ю. В.<sup>5</sup>  
(Учреждение Российской академии наук Институт прикладных  
математических исследований КарНЦ РАН, Петрозаводск)

*Рассматривается задача дележа затрат на создание общего сайта-коммуникатора для заданного целевого множества сайтов. В качестве математической модели предложена кооперативная игра, в которой игроками являются владельцы сайтов целевого множества, объединяющиеся для создания сайта-коммуникатора с целью уменьшения среднего количества кликов, требуемых для перехода с одного сайта на другой.*

Ключевые слова: вебометрика, веб-граф, кооперативная игра, дележ.

### **Введение**

Термин «вебометрика» (*webometrics*), введенный более 10 лет назад, обозначает раздел информатики, в рамках которого исследуются количественные аспекты конструирования и использования информационных ресурсов, структур и технологий примени-

<sup>1</sup> Работа выполнена при финансовой поддержке РФФИ, грант №08-07-00023-а.

<sup>2</sup> Владимир Викторович Мазалов, доктор физико-математических наук, профессор, ([vmazalov@krc.karelia.ru](mailto:vmazalov@krc.karelia.ru)).

<sup>3</sup> Андрей Анатольевич Печников, кандидат физико-математических наук, доцент, ([petchnikov@krc.karelia.ru](mailto:petchnikov@krc.karelia.ru)).

<sup>4</sup> Александр Владимирович Чирков, инженер-программист, ([avchirkov@krc.karelia.ru](mailto:avchirkov@krc.karelia.ru)).

<sup>5</sup> Юлия Васильевна Чуйко, кандидат физико-математических наук, ([julia@krc.karelia.ru](mailto:julia@krc.karelia.ru)).

тельно к *World Wide Web* (далее — Веб) [7]. К основным направлениям вебометрики можно отнести разработку средств сбора данных о Вебе, изучение гиперссылок, описательные и формальные модели Веба и социальные феномены в Вебе.

Предметом наших исследований является изучение структуры Веба и его фрагментов. Основным способом взаимодействия сайтов в Вебе, вследствие природы гипертекста, являются гиперссылки. Они образуют связи между сайтами, давая возможность пользователю переходить с одного ресурса на другой в поисках нужной информации. Множества сайтов и ссылок между ними могут быть представлены в виде графа, чем выше связность которого, тем доступнее представляемая веб-ресурсами информация. Для повышения связности интернет-графа кроме поисковых систем создаются различные каталоги сайтов, как общие с рубрикой (*list.mail.ru*, *www.yandex.ru*), так и тематические (списки научных, музыкальных, спортивных сайтов). При этом широко известные каталоги общего пользования содержат информацию о миллионах ресурсов и постоянно пополняются новыми, из-за чего фактически из таких каталогов пользователи попадают только на те ресурсы, что находятся на первых позициях их рейтингов. Поэтому часто для некоторого интернет-сообщества есть смысл в создании собственного ресурса, связывающего сайты всего сообщества.

В данной работе рассматривается задача дележа затрат на создание общего сайта-коммуникатора для заданного целевого множества сайтов. В качестве математической модели предложена кооперативная игра, в которой игроками являются владельцы сайтов целевого множества, объединяющиеся для создания сайта-коммуникатора с целью уменьшения среднего количества кликов, требуемых для перехода с одного сайта на другой. Здесь можно использовать различные классические принципы дележей, такие как нахождения ядер  $(C, K, N)$ , вектора Шепли, известные по работам Л. С. Шепли, О. Н. Бондаревой, М. Дэвиса, М. Машлера, И. Р. Д. Ауманна, Э. Мулена и более поздних исследователей [9, 1, 8, 6, 3]. Мы применяем и сравниваем три подхода: эго-

истичный пропорциональный раздел, вектор Шепли и модифицированный путем введения дополнительных поощрений вектор Шепли.

## 1. Постановка задачи

Целевым множеством  $T = \{t_1, t_2, \dots, t_n\}$  называется множество сайтов интернет-сообщества, сформированное по некоторому заданному признаку. Примерами целевых множеств могут служить сайты классических университетов России или сайты учреждений РАН.

Фрагментом Веба на множестве  $T$  называется множество сайтов  $T$  и множество всех гиперссылок, существующих между сайтами из  $T$ . Схемой фрагмента Веба называется ориентированный граф  $G = (T, E)$  без петель и кратных дуг, где  $E$  – множество дуг, причем дуга  $e(i, j) \in E$  существует тогда и только тогда, когда существует хотя бы одна гиперссылка с сайта, соответствующего вершине  $i$ , на сайт, соответствующий вершине  $j$ .

Дуги, т. е. гиперссылки между сайтами, могут различаться по весу (или «длине»). Действительно, при переходе между сайтами по гиперссылкам имеют большое значение такие характеристики «доступности» ссылки, как: глубина вложенности страницы, с которой сделана ссылка, относительно начальной страницы; сделана ссылка из меню или из сплошного текста на странице и т. п. Чем ссылка сильнее «спрятана», тем больше усилий нужно потратить на переход по ней, и тем большей считаем длину соответствующей ей дуги.

Одной из важных характеристик графа  $G = (T, E)$  является его связность. В работах [4, 5] показано, что важную роль в увеличении связности университетского и академического Веба играют веб-коммуникаторы, т. е. сайты, не входящие в целевое множество, но имеющие большое количество входящих ссылок с сайтов из  $T$  и/или исходящих ссылок на эти сайты. Как правило, веб-коммуникаторы возникают в Вебе «естественным образом», они не учитывают (да и не должны учитывать) специфику взаимосвязей целевого множества.

Предположим, что владельцы сайтов целевого множества реши-ли создать некий общий сайт, который в нашей терминологии называется веб-коммуникатором (например, ректоры университетов пришли к соглашению о создании сайта ассоциации уни-верситетов). Поскольку на создание и поддержку сайта требуются финансовые ресурсы, возникает вопрос о том, каким должен быть размер взноса каждого владельца сайта, входящего в  $T$ . Мы из-начально исходим из того, что в данном случае не используется принцип уравнивания, когда все участники платят одинаково. На самом деле ценность создаваемого веб-коммуникатора различна для разных участников, поэтому и взносы должны быть неодина-ковыми.

Может возникнуть ситуация, где создание коммуникатора не нужно ни одному из участников. Например, в сообществе офици-альных сайтов подразделений Карельского научного центра сайт Президиума КарНЦ РАН (<http://www.krc.karelia.ru>) имеет ссылки на все сайты институтов КарНЦ РАН, которые в свою очередь ссылаются на сайт Президиума. Здесь сайт Президиума уже фак-тически является таким коммуникатором, и создание дополни-тельного коммуникатора не уменьшит средний путь ни к одному из участников. Мы не будем рассматривать такие сообщества, считая, что создание коммуникатора выгодно хотя бы для одного из участников.

Дальнейшие рассуждения справедливы при условии, что граф  $G = (T, E)$  – сильно связный, т. е. существует путь ко-нечной длины, соединяющий любую пару вершин. Пусть  $d(i, t)$  – длина кратчайшего пути из вершины  $i$  в вершину  $t$  в графе  $G$ , где  $i, t \in T$ . Тогда критерием доступности сайта  $t$  на множестве  $T$  может быть средняя длина пути в заданную вершину  $t \in T$  из любой вершины  $i \in T, i \neq t$ , которая задается следующим образом:

$$midd(t) = \frac{\sum_{i \in T, i \neq t} d(i, t)}{n - 1}.$$

Предположим, что игроки договорились создать веб-коммуникатор  $h$ , с которого обязательно будут сделаны гиперс-

ссылки весом 1 на все сайты из  $T$ , и с каждого сайта из  $T$  будет сделана гиперссылка на  $h$ , также имеющая единичный вес. Тогда средняя длина пути в вершину  $t$  находится как

$$midd^h(t) = \frac{\sum_{i \in T, i \neq t} d^h(i, t)}{n - 1}.$$

В этом случае, с учетом путей, содержащих добавленную вершину  $h$ , для  $i \in T$  верно неравенство  $midd^h(i) \leq 2$  (либо на сайт  $i$  существуют ссылки с некоторых сайтов из  $T$  и длина пути равна 1, либо кратчайший путь проходит через сайт-коммуникатор  $h$  и длина пути равна 2). Тогда выигрыш  $i$ -го участника равен  $v(i) = midd(i) - midd^h(i)$ .

Обозначим  $Z$  стоимость разработки и поддержки сайта  $h$ , а взнос каждого игрока –  $z_i$ , причем  $Z = \sum_{i \in T} z_i$ . Требуется ответить на вопрос, каковы должны быть значения  $z_1, z_2, \dots, z_n$ , справедливые (в некотором смысле) для каждого игрока – владельца сайта целевого множества.

## 2. Эгоистичный подход

Один из простых подходов к разделению платы за создание коммуникатора – пропорциональный раздел. Считаем, что каждый участник действует исключительно в собственных интересах, принимая во внимание только ту выгоду, которую получит от участия в создании коммуникатора. Тогда соответствующим вариантом разделения будет назначение вклада каждого участника пропорционально выигрышу, который он получит.

*Пример 1.* Рассмотрим сообщество четырех сайтов, связанных между собой, как на рис. 1. Считаем, что дуга  $(0, 1)$  имеет длину 10, а все остальные – длину 1. Сайт 0 находится в самом выигрышном положении – на него с любого из сайтов сообщества можно попасть не более чем за два перехода длиной 1, в то время как он сам ставит ссылку только на сайт 1 и с очень большим весом, значительно увеличивая другим сайтам средние длины путей до них. Выигрыши участников от создания коммуникатора равны

соответственно  $\{0, 9, 6\frac{1}{3}, 3\}$ . Формально от создания коммуникатора сайт 0 ничего не выигрывает, следовательно, при данном подходе к разделению не должен вкладываться в создание нового сайта. Однако, это выглядит не совсем справедливым, учитывая, что сайт 0 ведет себя некорректно по отношению к остальным. •

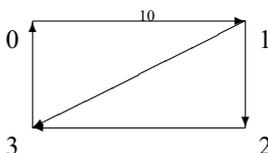


Рис. 1. Простое сообщество сайтов

Как видно из примера 1, при таком варианте разделения участники, которым создание коммуникатора формально не приносит никакой дополнительной выгоды, ничего не платят за создание коммуникатора.

### 3. Кооперативный подход

Рассматриваемая задача может быть также сформулирована с использованием методов кооперативной теории игр. Такой подход представляется более справедливым, так как разделение платы за создание коммуникатора выполняется уже с учетом поведения и ценности каждого из участников по отношению к коалиции. Пусть  $I = \{1, 2, \dots, n\}$  – множество игроков (владельцев сайтов  $t_1, t_2, \dots, t_n$ ), коалицией называется всякое подмножество  $S \in I$ . Мы считаем коалицией множество игроков, участвующих в создании коммуникатора, причем коммуникатор будет ссылаться только на участников коалиции, и только они будут ссылаться на коммуникатор. Обозначим  $v(S)$  выигрыш коалиции  $S$ , где  $v$  – характеристическая функция, которую требуется построить для изложенной выше содержательной постановки.

Выигрыш коалиции  $S$  равен  $v(S) = \sum_{i \in S} v(i)$ , где  $v(i) = \text{midd}(i) - \text{midd}_S^h(i)$  рассчитывается с учетом, того что коммуникатор создается только для членов коалиции  $S$ . Очевидно, что

для всех участников  $v(\{i\}) = 0$ , так как создание коммуникатора «только для себя» не изменяет длины ни одного из путей в веб-графе. Выигрыш гранд-коалиции – это суммарный выигрыш всех участников, когда коммуникатор создается для всех. Мы считаем, что в сообществе может создаваться только один коммуникатор, т. е. участники разделяются на тех, кто состоит в коалиции и участвует в создании коммуникатора, и тех, кто не участвует.

Предлагаемый подход основан на разделении платы за создание веб-коммуникатора пропорционально компонентам вектора Шепли, строящемуся с учетом среднего вклада каждого участника в образование выигрыша гранд-коалиции. Плата за создание сайта делится между членам сообщества  $T$  пропорционально величинам

$$\Phi_i = \sum_{S \subseteq T} \frac{(|S| - 1)!(n - |S|)!}{n!} (v(S) - v(S \setminus \{i\})),$$

равным долям каждого участника в общем выигрыше гранд-коалиции  $T$ .

*Пример 2.* Рассмотрим сообщество, описанное в примере 1. В результате кооперативного разделения игроки несут затраты на создание коммуникатора, пропорциональные компонентам вектора Шепли  $\{5\frac{1}{2}, 5\frac{2}{3}, 4\frac{1}{9}, 3\frac{1}{18}\}$ . Здесь значительную часть стоимости платит не только сайт, на который ведет самый «плохой» путь, но и сайт, который этот путь обеспечил. •

#### 4. Кооперативный подход с поощрениями

Подход, использованный в предыдущем примере, обеспечивает значительное финансовое участие в создании коммуникатора тех игроков, которые изначально создали плохие условия для остальных участников сообщества. Однако необходим способ стимуляции вступления таких игроков в коалицию, создающую коммуникатор. Предлагается следующая схема введения поощрений. Для каждого из игроков считается дополнительный

выигрыш, который он, вступая в коалиции, приносит игрокам, не участвующим в коалиции:

$$\Psi_i = \sum_{S \subseteq T: i \in S} \sum_{t \in T \setminus S} \left( \text{midd}_{S \setminus \{i\}}^h(t) - \text{midd}_S^h(t) \right).$$

Найдем отклонения данной суммы от среднего значения:  $\Delta_i = \Psi_i - \sum_{i \in T} \Psi_i / n$ , которые мы и будем использовать в качестве поощрений, отнимая их от значений компонент вектора Шепли. Сумма таких отклонений равна 0, часть из них отрицательны.

Участники сообщества  $T$  будут платить за создание сайта суммы, пропорциональные величинам  $\Phi_i - \Delta_i$ . Вопрос, могут ли эти значения быть отрицательными и в каких случаях, будет исследован в дальнейшей работе. В рамках данной модели вполне допускается отрицательность компонент дележа. Отрицательная плата для игрока означает, что участие его в коалиции настолько важно, в первую очередь тем, кто не в коалиции, что ему готовы заплатить.

*Пример 3.* Рассмотрим сообщество, в котором 3 сайта, 0 связан с 1 дугой длины 10, 1 связан с 2 и 2 связан с 0 дугами длины 1. При эгоистичном подходе плата за сайт-коммуникатор пропорциональна компонентам вектора  $\{0; 8,5; 4,5\}$ . При кооперативном получаем вектор  $\{4\frac{11}{12}, 4\frac{11}{12}, 3\frac{1}{6}\}$ . При кооперативном разделении с поощрениями  $\{3\frac{7}{12}, 3\frac{7}{12}, 5\frac{5}{6}\}$ . Здесь видно, что для последнего сайта плата за создание коммуникатора в схеме с поощрениями возрастает по сравнению с простым кооперативным разделением на основе вектора Шепли, его  $\Delta_2 = -2\frac{2}{3} < 0$ . За счет него поощряется вступление в коалицию первых двух участников, образующих пару, очень полезную для последнего участника в случае его невступления в коалицию. ●

## 5. Перспективы исследования

Для автоматизации проведения исследования и визуализации результатов было разработано приложение на языке *Java 2 SDK*, позволяющее для задаваемого сильно связанного веб-графа сооб-

щества находить варианты разделения платы за создание коммуникатора как с использованием эгоистичного, так и кооперативного подхода. Процедура расчета значений характеристических функций основана на модификации алгоритма Флойда [2] поиска путей в ориентированном графе. Для визуализации представления веб-графов использована *Java*-библиотека *Jung2 2.0* (<http://jung.sourceforge.net/>).

Данное приложение позволяет работать с множествами большой размерности и будет использовано для исследования таких целевых множеств, как российские научные сайты, университетские сайты и т. п.

### Литература

1. БОНДАРЕВА О. Н. *Некоторые применения методов линейного программирования к теории кооперативных игр* // Проблемы кибернетики. — 1963. — Т. 10. — С. 119–140.
2. КРИСТОФИДЕС Н. *Теория графов: алгоритмический подход*. — М.: Мир, 1978. — 429 с.
3. МУЛЕН Э. *Кооперативное принятие решений: Аксиомы и модели* Пер. с англ. — М.: Мир, 1991. — 464 с.
4. ПЕЧНИКОВ А. А. *О некоторых результатах вебометрических исследований университетского Веба* // IV Международная научно-практическая конференция «Современные информационные технологии и ИТ-образование». Сборник избранных трудов. — М.: ИНТУИТ.РУ. — 2009. — С. 530-537.
5. ЧУЙКО Ю. В., ПЕЧНИКОВ А. А. *Исследование связности российского научного Веба* // Труды Международной конференции «Когнитивный анализ и управление развитием ситуаций» (CASC'2009), 17-19 ноября 2009 г., Москва. — С. 283–286.
6. AUMANN R. J. *Some Non-Superadditive Games, and Their Shapley Values, in the Talmud* // International Journal of Game Theory. — 2010. — No. 39. — P. 3-10.

7. ALMIND T., INGWERSEN P. *Informetric analyses on the World Wide Web: Methodological approaches to “webometrics”* // Journal of Documentation. — 1997. — No. v 53(4). — P. 404-426.
8. DAVIS M., MASCHLER M. *The kernel of a cooperative game* // Naval Research Logistics Quarterly. — 1965. — Vol. 12. — P. 223–259.
9. SHAPLEY L. S. *On balanced sets and cores* // Naval Research Logistics Quarterly. — 1967. — T. 14. — C. 453–460.

## WEB-COMMUNICATOR CREATION COSTS SHARING PROBLEM AS A COOPERATIVE GAME

**Vladimir Mazalov**, Institute of Applied Mathematical Research of KarRC RAS, Petrozavodsk, Doctor of Science, professor (vmazalov@krc.karelia.ru).

**Andrey Pechnikov**, Institute of Applied Mathematical Research of KarRC RAS, Petrozavodsk, Cand.Sc., assistant professor (pechnikov@krc.karelia.ru).

**Alexandr Chirkov**, Institute of Applied Mathematical Research of KarRC RAS, Petrozavodsk, engineer (avchirkov@krc.karelia.ru).

**Julia Chuyko**, Institute of Applied Mathematical Research of KarRC RAS, Petrozavodsk, Cand.Sc. (julia@krc.karelia.ru).

*Abstract: We consider a problem of web-communicator creation costs sharing for a given target set of sites. We offer a cooperative model where agents are site owners which create a communicator site in purpose of reducing the average click count of surfing the sites.*

Keywords: webometrics, web-graph, cooperative game, utility allocation.

*Статья представлена к публикации членом редакционной коллегии М. В. Губко*