

РАНЖИРОВАНИЕ АКАДЕМИЧЕСКОГО ВЕБ-ПРОСТРАНСТВА

Мазалов В. В.¹, Печников А. А.²

*(Институт прикладных математических исследований
КарНЦ РАН, Петрозаводск)*

Трухина Л. И.³

*(Читинский институт (филиал) Байкальского государственного
университета экономики и права, Чита)*

Цынгугев Б. Т.⁴

(Забайкальский государственный университет, Чита)

Предлагается для анализа научно-образовательного веб-пространства использовать методы теории графов и кооперативной теории игр. Основной характеристикой анализа является центральность вершин во взвешенном графе. Взвешенный граф представлен как электрическая цепь, и в качестве меры центральности вершины рассматривается суммарный ток, протекающий через данную вершину. Вычисление центральности вершин использует обратную матрицу Лапласа. Рассматриваемый метод имеет относительно невысокую вычислительную сложность. Метод иллюстрируется результатами расчетов электрической центральности для ранжирования вершин графа, представляющего фрагмент взаимосвязанных

¹ Владимир Викторович Мазалов, доктор физико-математических наук, профессор (Петрозаводск, ул. Пушкина, д. 65, тел. (8142) 78-11-08, vmazalov@krc.karelia.ru)

² Андрей Анатольевич Печников, доктор технических наук, доцент (Петрозаводск, ул. Пушкинская, д. 11, тел. (8142) 76-63-12 (доп. 39), pechnikov@krc.karelia.ru)

³ Людмила Ивановна Трухина, кандидат физико-математических наук (litrukhina@gmail.ru).

⁴ Булат Тимурович Цынгугев, аспирант (btsynguev@gmail.com).

сайтов академических организаций РФ. Сделано сравнение с широко используемым методом PageRank.

Ключевые слова: электрическая центральность, вектор Майерсона, веб-граф

Введение

Анализ веб-пространства стремительно набирает всемирную популярность благодаря возникновению всемирной паутины Интернет, а также появлению в 1990-х годах онлайн-сервисов социальных сетей (SixDegrees, LiveJournal, Facebook, Twitter, YouTube и другие). Методы анализа социальных сетей чрезвычайно важны для анализа развития современных тенденций в различных областях науки, а также и для улучшения обмена научной информацией среди научных организаций.

Основной целью статьи является анализ академического веб-пространства России с использованием методов ранжирования вершин графов, основанных на обратной матрице Лапласа и на векторе Майерсона.

Одним из базовых понятий в анализе сетевых структур является *betweenness centrality*. Центральность вершины — это важная мера, отражающая то, насколько вершина участвует в процессе распространения информации между остальными вершинами в графе.

1. Электрическая центральность

В отличие от традиционных моделей вычисления центральности вершин в графе на основе геодезических путей или вычисления предельных вероятностей для некоторого случайного блуждания на данном графе (метод PageRank [6]) в работах [3, 5, 9–11, 13–15] было предложено при расчете меры центральности использовать аналогию с электрическими сетями. В данном подходе граф рассматривается как электрическая цепь с идеальными элементами, где каждое ребро имеет некую пропускную способность (значение обратное сопротивлению), а верши-

ны графа являются её узлами. Для поиска меры центральности в модели электрической цепи используются правила Кирхгофа. Одну модель отличает от другой то, что напряжение подается в цепь различными способами и в качестве меры центральности вершины выбирается либо ее электрический потенциал, либо ток, проходящий через данную вершину, либо другие характеристики.

В [4] рассматривается модель электрической цепи, где каждая вершина цепи соединена с искусственно вводимой вершиной v_{n+1} ребром с пропускной способностью δ . Единичный электрический ток подается в некоторую вершину s , цепь заземляется в вершине v_{n+1} . Таким образом, электрический ток течет по цепи из вершины s в вершину v_{n+1} . Подобный прием использовался в работах [7, 8].

Мерой центральности вершины v служит средняя величина тока, проходящего через вершину v по всем возможным s . Отметим, что ток на ребрах, инцидентных вершине v_{n+1} , не учитывается при расчете центральности.

Обозначим взвешенный граф как $G = (V, E, W)$, где V — множество вершин, E — множество ребер, W — матрица весов:

$$W(G) = \begin{pmatrix} 0 & w_{1,2} & \dots & w_{1,n} \\ w_{2,1} & 0 & \dots & w_{2,n} \\ \vdots & \vdots & \ddots & \vdots \\ w_{n,1} & w_{n,2} & \dots & 0 \end{pmatrix},$$

где $w_{i,j} \geq 0$ — вес ребра между вершинами v_i и v_j , $n = |V|$ — число вершин. Если вершины v_i и v_j несмежные, то $w_{i,j} = 0$. Если G — неориентированный граф, то $w_{i,j} = w_{j,i}$.

Пусть φ_v^s обозначает абсолютный потенциал в вершине $v \in V$ при условии, что источник электрического тока подключен к вершине s . Примем абсолютный потенциал в вершине v_{n+1} равным нулю: $\varphi_{v_{n+1}}^s = 0$, тогда вектор абсолютных потенциалов $\varphi^s(G) = [\varphi_{v_1}^s, \dots, \varphi_{v_n}^s]^T$ в вершинах графа G можно определить следующим выражением (правила Кирхгофа) [4]:

$$(1) \quad \varphi^s(G) = [D(G) - W(G) + \delta I]^{-1} b_s,$$

где $D(G)$ — диагональная матрица суммы весов ребер, инцидентных соответствующим вершинам, I — единичная матрица размерности n , b_s — вектор-столбец, состоящий из n элементов, значения которых равны:

$$b_s(v) = \begin{cases} 1, & v = s, \\ 0 & \text{иначе.} \end{cases}$$

Ток, протекающий через ребро $e = (v_i, v_j)$, согласно закону Ома, равен $x_e^s = |\varphi_{v_i}^s - \varphi_{v_j}^s| \cdot w_{i,j}$.

Величину тока, протекающего через вершину v можно найти, определив сумму токов на ребрах, инцидентных ей, и поделив полученную сумму на 2. Отметим, что необходимость в делении на 2 возникла в связи с тем, что при суммировании всех токов на ребрах, инцидентных вершине, учитывается электрический ток как входящий в вершину, так и выходящий из нее.

Соответственно величину тока, протекающего через вершину v , при условии, что источник тока находится в вершине s , можно вычислить следующим образом:

$$(2) \quad x^s(v) = \frac{1}{2}(b_s(v) + \sum_{e:v \in e} x_e^s).$$

Таким образом, меру центральности в вершине v для взвешенных графов $CF_\delta(v)$ можно определить по формуле:

$$(3) \quad CF_\delta(v) = \frac{1}{n} \sum_{s \in V} x^s(v) = \frac{1}{2n} (1 + \sum_{s \in V} \sum_{e:v \in e} x_e^s).$$

Вычислительная сложность алгоритма поиска электрической центральности вершин взвешенных графов предлагаемым способом относительно невысока и равна сложности вычисления обратной матрицы $O(n^3)$.

2. Вектор Майерсона

Для определения центральности вершин в графе можно использовать теоретико-игровой подход.

Пусть множество вершин графа G является множеством игроков кооперативной игры, в которой характеристическая функция задаётся следующим образом [12]: каждая прямая связь — путь длиной 1 — приносит игрокам доход r , где $0 \leq r \leq 1$. Кроме того, игроки также извлекают выгоду из косвенных (непрямых) связей, но уже меньшую. За каждый путь длиной 2 коалиция получает r^2 , за путь длиной 3 получает r^3 и т.д. Так как любые две вершины могут быть связаны несколькими путями разной длины, то при построении характеристической функции учитываются только кратчайшие. А если две вершины связаны несколькими путями одинаковой длины, то учитываются все такие пути, при условии, что между этими вершинами нет другого меньшего пути.

Таким образом, для любой коалиции S можно записать

$$(4) \quad \vartheta(S) = a_1 r + a_2 r^2 + \dots + a_L r^L = \sum_{k=1}^L a_k r^k,$$

где L — максимальное расстояние между двумя вершинами в данной коалиции;

a_k — число геодезических путей длины k в данной коалиции,

$$\vartheta(i) = 0, \quad \forall i \in N.$$

Делёж в данной игре определяется по формуле

$$(5) \quad Y_i(\vartheta, G) = \frac{A_1^i}{2} r + \frac{A_2^i}{3} r^2 + \dots + \frac{A_L^i}{L+1} r^L = \sum_{k=1}^L \frac{A_k^i}{k+1} r^k,$$

где A_k^i — число путей длины k , содержащих игрока i .

В [2] доказано, что предложенный делёж для невзвешенных графов является вектором Майерсона.

В рассматриваемой игре значение вектора Майерсона зависит от числа путей. Следовательно, большее значение выигрыша для игрока означает большее количество путей, проходящих через него, а значит, и больший вес игрока в сети. Поэтому можно использовать вектор Майерсона как меру центральности вершин графа. Модифицируем этот подход для взвешенных сетей.

Пусть дан взвешенный граф с целыми весами на ребрах. Заменим каждое ребро веса n на n параллельных ребер единицы веса. Самый короткий путь между любыми двумя вершинами будет определяться так же, как и в невзвешенном графе, так как все ребра имеют единичную длину. Но теперь в целом геодезических путей больше, чем было ранее, из-за мультиребер. Если вершины A и B соединены m ребрами, а B и C — n ребрами, то из вершины A в вершину C ведёт $m \cdot n$ путей. Применяя формулу (5) к вершинам мультиграфа, получим меру центральности вершины во взвешенном графе.

В [2] предложен алгоритм нахождения числа кратчайших путей в невзвешенном графе с помощью производящих функций. Этот алгоритм несложно перенести на случай взвешенных графов.

Рассмотрим дерево G_p с корнем в вершине p . Введём в рассмотрение производящую функцию

$$\varphi_p(x) = \sum_{k=1}^L \alpha_k^p x^k,$$

где α_k^p — число путей, состоящих из k игроков (длины $k - 1$), содержащих вершину p .

Вычислим производящую функцию рекуррентно. Вначале определим в финальных вершинах q дерева G_p

$$\varphi_q(x) = x.$$

Для остальных вершин дерева, не являющихся корнем, положим

$$(6) \quad \varphi_q(x) = x \left(1 + \sum_{i=1}^d \varphi_{q_i}(x) \right),$$

где сумма берётся по всем потомкам $q_i, i = 1, \dots, d$, вершины q .

Для корня дерева — вершины p — положим

$$(7) \quad \varphi_p(x) = x \left(1 + \sum_{i=1}^d \varphi_{q_i}(x) + \sum_{i \neq j} \varphi_{q_i}(x) \varphi_{q_j}(x) \right),$$

где сумма берётся по всем потомкам $q_i, i = 1, \dots, d$, вершины p .

Во взвешенном графе число путей через производящую функцию вычисляется точно так же, с той лишь разницей, что когда поднимаемся вверх к корню, то в вершине, где есть параллельные ребра, мы как бы копируем поддерево внизу такое же число раз. То есть вычисляем во всех вершинах $\varphi_q(x) = x(1 + \sum_i w(q, q_i)\varphi_{q_i}(x))$, где, q_i - потомки q и $w(q, q_i)$ – вес ребра (q, q_i) . А в корне функция вычисляется по формуле

$$(8) \quad \varphi_p(x) = x \left(1 + \sum_{i=1}^d w(p, q_i)\varphi_{q_i}(x) + \sum_{i \neq j} w(p, q_i)\varphi_{q_i}(x) \cdot w(p, q_j)\varphi_{q_j}(x) \right).$$

3. Примеры

Рассмотрим частный случай, когда G – полный двудольный граф с n вершинами, вершины которого разделены на два множества V_1 и V_2 таким образом, что ни одна вершина в V_1 не соединена с вершинами V_1 , и ни одна вершина в V_2 не соединена с вершинами V_2 . Обозначим данный полный двудольный граф как $K_{|V_1|, |V_2|}$. Все ребра имеют вес равный единице. Для случая $K_{1, n-1}$, т.е. для звезды, аналитическое решение электрической центральности было представлено в [4].

3.1. ДВУДОЛЬНЫЙ ГРАФ $K_{2, N-2}$.

Пусть $V_1 = \{v_1, v_2\}$, $v' \in V_2$. Тогда матрица Лапласа имеет вид:

$$D(G) - W(G) + \delta I = \begin{pmatrix} n-2+\delta & 0 & -1 & -1 & \dots & -1 \\ 0 & n-2+\delta & -1 & -1 & \dots & -1 \\ -1 & -1 & 2+\delta & 0 & \dots & 0 \\ -1 & -1 & 0 & 2+\delta & \dots & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ -1 & -1 & 0 & 0 & \dots & 2+\delta \end{pmatrix}.$$

Обратная матрица имеет вид:

$$(D(G) - W(G) + \delta I)^{-1} = \frac{1}{\delta(n + \delta)} \times$$

$$\times \begin{pmatrix} \frac{n-2+n\delta+\delta^2}{n+\delta-2} & \frac{n-2}{n+\delta-2} & 1 & 1 & \dots & 1 \\ \frac{n-2}{n+\delta-2} & \frac{n-2+n\delta+\delta^2}{n+\delta-2} & 1 & 1 & \dots & 1 \\ 1 & 1 & \frac{2+n\delta+\delta^2}{2+\delta} & \frac{2}{2+\delta} & \dots & \frac{2}{2+\delta} \\ 1 & 1 & \frac{2}{2+\delta} & \frac{2+n\delta+\delta^2}{2+\delta} & \dots & \frac{2}{2+\delta} \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & 1 & \frac{2}{2+\delta} & \frac{2}{2+\delta} & \dots & \frac{2+n\delta+\delta^2}{2+\delta} \end{pmatrix}.$$

Здесь и далее, тот факт что это обратная матрица, проверяется непосредственно умножением на матрицу $D(G) - W(G) + \delta I$. После несложных вычислений приходим к единичной матрице.

При $s = v_1$ ток распределяется следующим образом:

$$x^s(v_1) = \frac{1}{2} \left(1 + \frac{(\delta + n - 1)(n - 2)}{(n + \delta)(n + \delta - 2)} \right),$$

$$x^s(v_2) = \frac{(n - 2)}{2(n + \delta - 2)(\delta + n)},$$

$$x^s(v') = \frac{1}{2(n + \delta - 2)}.$$

Так как вершины v_1 и v_2 расположены симметрично, то при $s = v_2$ получаем

$$x^s(v_1) = \frac{(n - 2)}{2(n + \delta - 2)(\delta + n)},$$

$$x^s(v_2) = \frac{1}{2} \left(1 + \frac{(\delta + n - 1)(n - 2)}{(n + \delta)(n + \delta - 2)} \right),$$

$$x^s(v') = \frac{1}{2(n + \delta - 2)}.$$

При $s = v'$:

$$x^s(v_1) = x^s(v_2) = \frac{\delta + 2n - 4}{2(2 + \delta)(\delta + n)},$$

$$x^s(s) = \frac{1}{2} \left(1 + \frac{2(\delta + n - 1)}{(2 + \delta)(\delta + n)} \right),$$

$$x^s(v') = \frac{1}{(2 + \delta)(\delta + n)}.$$

Получаем следующий результат для $K_{2,n-2}$:

$$CF_\delta(v_1) = CF_\delta(v_2) = \frac{1}{2n} \left(1 + \frac{n-2}{n+\delta-2} + \frac{(n-2)(\delta+2n-4)}{(2+\delta)(\delta+n)} \right),$$

$$CF_\delta(v') = \frac{1}{2n} \left(1 + \frac{2}{n+\delta-2} + \frac{2(\delta+2n-4)}{(2+\delta)(\delta+n)} \right).$$

3.2. ДВУДОЛЬНЫЙ ГРАФ $K_{3,N-3}$.

Пусть $V_1 = \{v_1, v_2, v_3\}$, $v' \in V_2$.

Тогда матрица Лапласа:

$$D(G) - W(G) + \delta I =$$

$$= \begin{pmatrix} n-3+\delta & 0 & 0 & -1 & -1 & \dots & -1 \\ 0 & n-3+\delta & 0 & -1 & -1 & \dots & -1 \\ 0 & 0 & n-3+\delta & -1 & -1 & \dots & -1 \\ -1 & -1 & -1 & 3+\delta & 0 & \dots & 0 \\ -1 & -1 & -1 & 0 & 3+\delta & \dots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ -1 & -1 & -1 & 0 & 0 & \dots & 3+\delta \end{pmatrix}.$$

Обратная матрица имеет вид:

$$(D(G) - W(G) + \delta I)^{-1} = \frac{1}{\delta(n+\delta)} \times$$

$$\times \begin{pmatrix} \frac{n-3+n\delta+\delta^2}{n+\delta-3} & \frac{n-3}{n+\delta-3} & \frac{n-3}{n+\delta-3} & 1 & 1 & \dots & 1 \\ \frac{n-3}{n+\delta-3} & \frac{n-3+n\delta+\delta^2}{n+\delta-3} & \frac{n-3}{n+\delta-3} & 1 & 1 & \dots & 1 \\ \frac{n-3}{n+\delta-3} & \frac{n-3}{n+\delta-3} & \frac{n-3+n\delta+\delta^2}{n+\delta-3} & 1 & 1 & \dots & 1 \\ 1 & 1 & 1 & \frac{3+n\delta+\delta^2}{3+\delta} & \frac{3}{3+\delta} & \dots & \frac{3}{3+\delta} \\ 1 & 1 & 1 & \frac{3}{3+\delta} & \frac{3+n\delta+\delta^2}{3+\delta} & \dots & \frac{3}{3+\delta} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & 1 & 1 & \frac{3}{3+\delta} & \frac{3}{3+\delta} & \dots & \frac{3+n\delta+\delta^2}{3+\delta} \end{pmatrix}$$

Рассуждая аналогично $K_{2,n-2}$, получим следующий результат для $K_{3,n-3}$:

$$\begin{aligned} CF_{\delta}(v_1) &= CF_{\delta}(v_2) = CF_{\delta}(v_3) = \\ &= \frac{1}{2n} \left(1 + \frac{(n-3)(\delta+n+1)}{(n+\delta-3)(\delta+n)} + \frac{(n-3)(\delta+2n-5)}{(3+\delta)(\delta+n)} \right), \\ CF_{\delta}(v') &= \frac{1}{2n} \left(1 + \frac{3(\delta+n+1)}{(n+\delta-3)(\delta+n)} + \frac{3(\delta+2n-5)}{(3+\delta)(\delta+n)} \right). \end{aligned}$$

3.3. ДВУДОЛЬНЫЙ ГРАФ $K_{R,N-R}$.

Пусть $v \in V_1$, $v' \in V_2$ и $r = |V_1|$, $n-r = |V_2|$. Подобно рассуждениям рассмотренных случаев $K_{1,n-1}$, $K_{2,n-2}$ и $K_{3,n-3}$, можно определить значения центральностей для вершин полного двудольного графа $K_{r,n-r}$:

$$\begin{aligned} CF_{\delta}(v) &= \frac{1}{2n} \left(1 + \frac{(n-r)(\delta+n-2+r)}{(n+\delta-r)(\delta+n)} + \right. \\ &\quad \left. + \frac{(n-r)(\delta+2n-2-r)}{(r+\delta)(\delta+n)} \right), \\ CF_{\delta}(v') &= \frac{1}{2n} \left(1 + \frac{r(\delta+n-2+r)}{(n+\delta-r)(\delta+n)} + \frac{r(\delta+2n-2-r)}{(r+\delta)(\delta+n)} \right). \end{aligned}$$

3.4. ЦЕНТРАЛЬНОСТЬ НА ОСНОВЕ ВЕКТОРА МАЙЕРСОНА ДЛЯ ЗВЕЗДЫ $K_{1,N-1}$.

Найдём меру центральности на основе вектора Майерсона для звезды из n вершин. Пусть вес ребра между вершинами 1 и 2 равен $w_{1,2}$, а остальные рёбра имеют вес равный единице.

По формуле (5) для центра звезды получим

$$Y_1 = \frac{A_1^1}{2}r + \frac{A_2^1}{3}r^2 = \frac{n-2+w_{1,2}}{2}r + \left(\frac{C_{n-2}^2 + w_{1,2}(n-2)}{3} \right) r^2 = \\ \frac{n-2+w_{1,2}}{2}r + \left(\frac{(n-2)(n-3)}{6} + \frac{w_{1,2}(n-2)}{3} \right) r^2.$$

Для вершины 2:

$$Y_2 = \frac{A_1^2}{2}r + \frac{A_2^2}{3}r^2 = \frac{w_{1,2}}{2}r + \frac{w_{1,2}(n-1)}{3}r^2.$$

Для всех остальных вершин:

$$Y_v = \frac{A_1^v}{2}r + \frac{A_2^v}{3}r^2 = \frac{1}{2}r + \frac{n-2+w_{1,2}}{3}r^2.$$

4. Эксперименты

В качестве примера был рассмотрен фрагмент реального веб-графа научных учреждений Российской академии наук (далее – веб-граф РАН) в их дореформенной версии, построенного с использованием базы данных внешних гиперссылок [1]. База данных расположена в интернете по адресу <http://grid.krc.karelia.ru/webometrics2>. Для гостевого входа в качестве имени пользователя и пароля достаточно ввести слово «guest».

Веб-граф РАН представляет собой ориентированный граф с кратными дугами без петель. По данным на март 2014 года он содержал 956 вершин, соответствующих сайтам научных отделений, центров, институтов, библиотек и др., связанных почти 39000 дугами, соответствующими гиперссылкам, связывающим эти сайты.

Поскольку в статье рассматривается неориентированный граф, для проведения вычислительных экспериментов над веб-графом РАН были выполнены следующие действия:

- 1) оставлены только те вершины веб-графа, любая пара которых связана «встречными» дугами любой кратности;
- 2) все кратные встречные дуги заменены на ребра, вес которых равен меньшей из двух кратностей встречных дуг;
- 3) в полученном после первых двух шагов графе удалены все вершины (и соответствующие им ребра), не вошедшие в максимальную компоненту связности.

Далее из построенного графа была удалена вершина, соответствующая сайту РАН с доменным именем www.ras.ru. Дело в том, что сайт РАН занимает доминирующее положение в системе веб-сайтов учреждений РАН, имея около 500 исходящих гиперссылок и 1300 входящих. Вследствие этого при любом методе ранжирования вершин веб-графа, вершина, соответствующая данному сайту, занимает первое место в рейтингах. Поскольку наша цель – показать, что новые предлагаемые методы ранжирования вершин обладают более высокой чувствительностью по отношению к вершинам с определенными особенностями, доминирующую вершину мы исключили из дальнейшего рассмотрения.

Полученный таким образом граф содержит 169 вершин, связанных 279 ребрами. Вес ребер варьируется от 1 до 38, средний вес ребра равен 2,04. Его изображение приведено на рис. 1.

На рисунке некоторые вершины графа помечены названиями научных учреждений. Пронумерованными эллипсами выделены четко различимые группы веб-сайтов, таких как:

1–3. Сайты научных учреждений, входящих в состав Дальневосточного, Сибирского и Уральского региональных отделений РАН (ДВО РАН, СО РАН и УрО РАН).

4. Сайты научных учреждений, входящих в состав КарНЦ РАН.

5–6. Сайты институтов (Институт социально-экономического развития территорий РАН, Институт теоретической физики РАН).

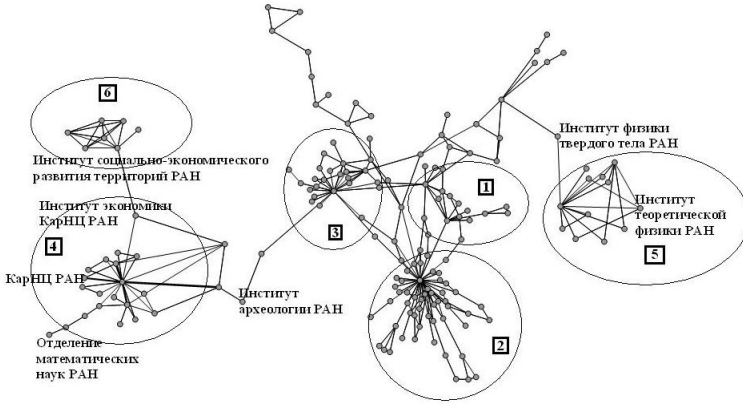


Рис. 1. Неориентированный граф учреждений РАН

Достаточно очевидно, что сайты, являющиеся «головными» в этих группах, должны иметь высокие значения такой известной характеристики ссылочного ранжирования как PageRank (PR), введенной в работе [6]. И действительно, это подтверждается таблицей 1, в которую попали все сайты учреждений, указанных в пп. 1–5.

Таблица 1. Десять первых сайтов с наибольшим значением PR

Название	URL	PR
Сибирское отделение РАН	www.sbras.nsc.ru	0,088
Карельский научный центр РАН	www.krc.karelia.ru	0,048
Институт вычислительных технологий СО РАН	www.ict.nsc.ru	0,038
Уральское отделение РАН	www.uran.ru	0,037
Институт теоретической физики им. Л.Д. Ландау РАН	itp.ac.ru	0,022
Дальневосточный геологический институт ДВО РАН	www.fegi.ru	0,017
Дальневосточное отделение РАН	www.febras.ru	0,013
Институт социально-экономического развития территорий РАН	www.vscac.ru	0,013
Институт проблем химической физики РАН	www.icp.ac.ru	0,012
Институт прикладных математических исследований КарНЦ РАН	mathem.krc.karelia.ru	0,012

Для этих же сайтов были сделаны расчеты значений электрической центральности и центральности по вектору Майерсона. Сравнивая значения PR с ранжированием по значениям электрической центральности для $\delta = 0,3$, можно отметить, что 5 из 6 указанных головных сайтов попадают в первую десятку и в этом случае, а сайт www.vssc.ac.ru находится на 15-м месте.

Несколько иные результаты мы имеем для вектора Майерсона при $r = 0,9$: только 3 головных сайта попадают в первую десятку.

Таблица 2. Сравнение рангов 5 учреждений РАН

Название (URL)	Ранг по PR	Ранг по электрической центральности	Ранг по вектору Майерсона
Институт археологии РАН (www.archaeolog.ru)	76	41	3
Институт истории и археологии УрО РАН (www.ihist.uran.ru)	133	38	5
Институт экономики КарНЦ РАН (economy.krc.karelia.ru)	80	17	7
Библиотека РАН (www.rasl.ru)	151	19	35
Институт физики твердого тела РАН (www.issp.ac.ru)	137	23	22

На рис. 1 несложно выявить сайты, удаление которых ведет к образованию достаточно крупных несвязных компонент. В первую очередь, к ним относятся сайты Института археологии РАН, Института истории и археологии УрО РАН, Института экономики КарНЦ РАН, Библиотеки РАН и Института физики твердого тела РАН. Значения рангов сайтов этих учреждений по PR, электрической центральности ($\delta = 0,3$) и вектору Майерсона ($r = 0,9$) приводятся в таблице 2. Здесь в колонке «Ранг по PR» даются порядковые номера указанных сайтов в общем списке сайтов, упорядоченных по убыванию значений PR. Такой же смысл имеют значения, указанные в колонках «Ранг по электри-

ческой центральности» и «Ранг по вектору Майерсона».

Очевидно, что обнаружение указанных сайтов, имеющих важное значение в анализируемом графе РАН, более точно проводится с помощью методов, основанных на использовании понятий электрической центральности и вектора Майерсона, нежели с помощью PR. Например, сайт Института физики твердого тела РАН по сравнению с рангом по PR (137 место) поднялся на 115 позиций в ранге по вектору Майерсона (22 место). На рис. 1 видно, что удаление этого сайта ведет к образованию двух несвязных компонент, к одной из которых относится группа, состоящая из 12 сайтов, входящих в веб-пространство Института теоретической физики РАН, включающего официальный сайт, сайты лабораторий, проектов, библиотеки и др.

5. Заключение

В работе предложен подход к анализу академического веб-пространства на основе разработанных авторами методов вычисления центральности вершин во взвешенном графе. Граф моделируется как некоторая электрическая цепь с новой заземленной вершиной, затем с одинаковой вероятностью в каждую из вершин подводится единичный ток. Средняя величина тока, проходящего через вершину, и берется в качестве центральности данной вершины. В физическом смысле это соответствует второму началу термодинамики: система находится в термодинамическом равновесии, и энтропия такой равновесной системы является мерой центральности в сети. Информационная энтропия тесно связана с термодинамической энтропией. В этом смысле распространение информации в академической сети подобно распространению тока в электрической цепи. Для вычисления центральности предлагается использовать правила Кирхгофа, которые представляют собой систему линейных уравнений. Для некоторых видов графов центральность можно найти в аналитическом виде. В работе это сделано для двудольных графов. В общем случае задача сводится к вычислению обратной матрицы, для этого можно использовать численные методы. Предложенный метод был исполь-

132

зован для моделирования академического веб-пространства, сделано сравнение с традиционным методом PageRank и вектором Майерсона, адаптированным для взвешенных графов. Показаны преимущества метода электрической центральности на примере вычисления этой характеристики для вершин, которые являются коммуникаторами связанных компонент.

Работа частично поддержана Отделением математических наук РАН, грантом РГНФ (проект 15-02-00352) и грантами РФФИ (проекты 15-01-06105, 16-51-55006).

Литература

1. ГОЛОВИН А.С., ПЕЧНИКОВ А.А. *База данных внешних гиперссылок для исследования фрагментов Веба* // Информационная среда вуза XXI века: материалы VII Всероссийской научно-практической конференции, Петрозаводск, 23–27 сентября 2013 г. – С. 55–57.
2. МАЗАЛОВ В.В., ТРУХИНА Л.И. *Производящие функции и вектор Майерсона в коммуникационных сетях* // Дискретная математика. – 2014. – Том 26, Вып. 3. – С. 65–75.
3. ЧЕБОТАРЕВ П.Ю., ШАМИС Е.В. *Матричная теорема о лесах и измерение связей в малых социальных группах* // Автоматика и телемеханика. – 1997. – №9. – С. 125–137.
4. AVRACHENKOV K.E., MAZALOV V.V., TSYNGUEV B.T. *Beta Current Flow Centrality for Weighted Networks* // Lecture Notes in Computer Science. Computational Social Networks. – 2015. – Vol. 9197. – P. 216–227. (DOI: 10.1007/978-3-319-21786-4_19).
5. BRANDES U., FLEISCHER D. *Centrality measures based on current flow* // STACS-2005. Lecture Notes in Computer Science. – 2005. – Vol. 3404. – P. 533–544. (DOI:10.1007/978-3-540-31856-9_44).
6. BRIN S., PAGE L. *The anatomy of a large-scale hypertextual Web search engine* // Computer Networks and ISDN Systems. – 1998. – Vol. 30(17). – P. 107–117. (DOI: 10.1016/j.comnet.2012.10.007).

7. CHEBOTAREV P., SHAMIS E. *The Forest Metrics for Graph Vertices* // *Electronic Notes in Discrete Mathematics*. – 2002. – Vol. 11. – P. 98–107.
8. CHEBOTAREV P., AGAEV R. *Forest matrices around the Laplacian matrix* // *Linear Algebra and its Applications*. – 2002. – Vol. 356, Issues 1–3. – P. 253–274.
9. ESTRADA E., HATANO N. *Resistance Distance, Information Centrality, Node Vulnerability and Vibrations in Complex Networks* // *Network Science. Complexity in Nature and Technology*, Springer, 2010. – P. 13–29. (DOI:10.1007/978-1-84996-396-1_2).
10. FREEMAN L.C., BORGATTI S.P., WHITE D.R. *Centrality in valued graphs: A measure of betweenness based on network flow* // *Social Networks* 13. – 1991. – P. 141–154. (DOI:10.1016/0378-8733(91)90017-N).
11. HINES P., BLUMSACK S. *A centrality measure for electrical networks* // *Proc. 41st Hawaii International Conference on System Sciences*. – 2008. – P. 185. (DOI: 10.1109/HICSS.2008.5).
12. JACKSON M.O. *Social and economic networks*. – Princeton University Press, 2008.
13. NEWMAN M.E.J. *A measure of betweenness centrality based on random walks* // *Social networks*. – 2005. – Vol. 27. – P. 39–54. (DOI:10.1016/j.socnet.2004.11.009).
14. WANG H., HUA H., WANG D. *Cacti with minimum, second-minimum, and third-minimum Kirchhoff indices* // *Mathematical Communications*. – Vol. 15. – 2010. – P. 347–358.
15. WANG Z., SCAGLIONE A., THOMAS R. *Electrical centrality measures for electric power grid vulnerability analysis* // *Proc. 49th IEEE Conference on Decision and Control (CDC)*. – 2010. – P. 5792–5797. (DOI: 10.1109/CDC.2010.5717964).

RANKING OF RUSSIAN ACADEMIC WEB

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

Andrey Pechnikov, Institute of Applied Mathematical Research Karelian Research Centre RAS, Petrozavodsk, Doctor of Science, associate professor (pechnikov@krc.karelia.ru).

Liudmila Trukhina, Chita Branch of Baikal State University of Economics and Law, Chita, (litrukhdina@gmail.ru).

Bulat Tsynguev, Transbaikal State University, Chita, (btsynguev@gmail.com).

Abstract: We propose to apply the methods of the graph theory and the cooperative game theory for determining nodes centrality in weighted graphs. In this method a graph considers like electric circuit, where the centrality measure is the total sum of electric current flow through this node. The computing of betweenness centrality is based on the inverse Kirchhoff's matrix. This method has low computational complexity. The results of numerical experiments of electric centrality are presented for graph of web sites of Russian academic organizations. We compare our results with the popular Page Rank method.

Keywords: Electric centrality, Myerson value, webgraph.

Статья представлена к публикации членом редакционной коллегии П.Ю. Чеботарёвым.

Поступила в редакцию 25.08.2015.

Дата опубликования 31.05.2016.