

¹ Институт математики им. С. Л. Соболева СО РАН
пр. Акад. Коптюга, 4, Новосибирск, 630090, Россия

² Новосибирский государственный университет
ул. Пирогова, 2, Новосибирск, 630090, Россия

E-mail: vesnin@math.nsc.ru; e_konsta@math.nsc.ru
savinmike.u@gmail.com

О СЦЕНАРИЯХ ПРИСОЕДИНЕНИЯ НОВЫХ САЙТОВ К ВЕБ-ПРОСТРАНСТВУ СО РАН*

Рассматриваются преобразования веб-графа Сибирского отделения Российской академии наук. Вершины графа соответствуют сайтам научных организаций; две вершины соединяются ориентированным ребром, если с одного сайта на другой имеется ссылка. В работе моделируются сценарии присоединения к этому графу новых вершин и исследуются происходящие в результате этого изменения рангов вершин, вычисляемых алгоритмом PageRank. Находятся значения структурных параметров графов, гарантирующие выход новых вершин в лидеры.

Ключевые слова: веб-граф, веб-пространство, PageRank, ранг вершины, сайт.

Введение

Предметом изучения в данной работе является веб-граф научного пространства Сибирского отделения Российской академии наук. Под веб-графом научного пространства мы понимаем ориентированный граф, вершины которого соответствуют сайтам научных учреждений, а дуги – их ссылкам друг на друга. Веб-граф научного пространства СО РАН был впервые построен в [1]. Очевидно, что с течением времени веб-граф может меняться. В данной работе рассматривается веб-граф научного пространства СО РАН по состоянию на 2 ноября 2012 г., представленный на сайте объединенного ученого совета СО РАН по нанотехнологиям и информационным технологиям¹. Вебометрический анализ и граф-теоретический анализ веб-пространства СО РАН представлен в [2].

Популярным и хорошо зарекомендовавшим себя средством ранжирования сайтов веб-пространства является алгоритм PageRank [3]. Он также активно используется при изучении связей графов научного сотрудничества и структуры веб-графов (см. [4; 5]). Основанный на алгоритме PageRank анализ научного веб-пространства СО РАН реализован в [6]. Поскольку веб-пространство развивается и меняется с течением времени, его мониторинг представляет несомненный интерес. Изменения веб-пространства обусловлены появлением новых сайтов, установлением новых информационных связей, утратой существовавших связей. В [7] с помощью алгоритма PageRank были проведены ранжирования сайтов веб-пространства СО

* Работа выполнена при финансовой поддержке Президиума СО РАН (междисциплинарный интеграционный проект № 21, 2012–2014) и РФФИ (проект № 12-01-00448).

¹ См.: <http://ousnano.sbras.ru/sitepage.php?PageID=2128>

РАН по состоянию на 2 ноября 2012 г. Соответствующий веб-граф G_{106} имеет 106 вершин и 1 084 дуги. Распределение рангов вершин в данном графе приведено на рис. 1. Наибольший ранг, равный 0,079553, имеет вершина, соответствующая сайту Портал СО РАН. Информация о рангах первых 80 вершин графа G_{106} приведена в [7].

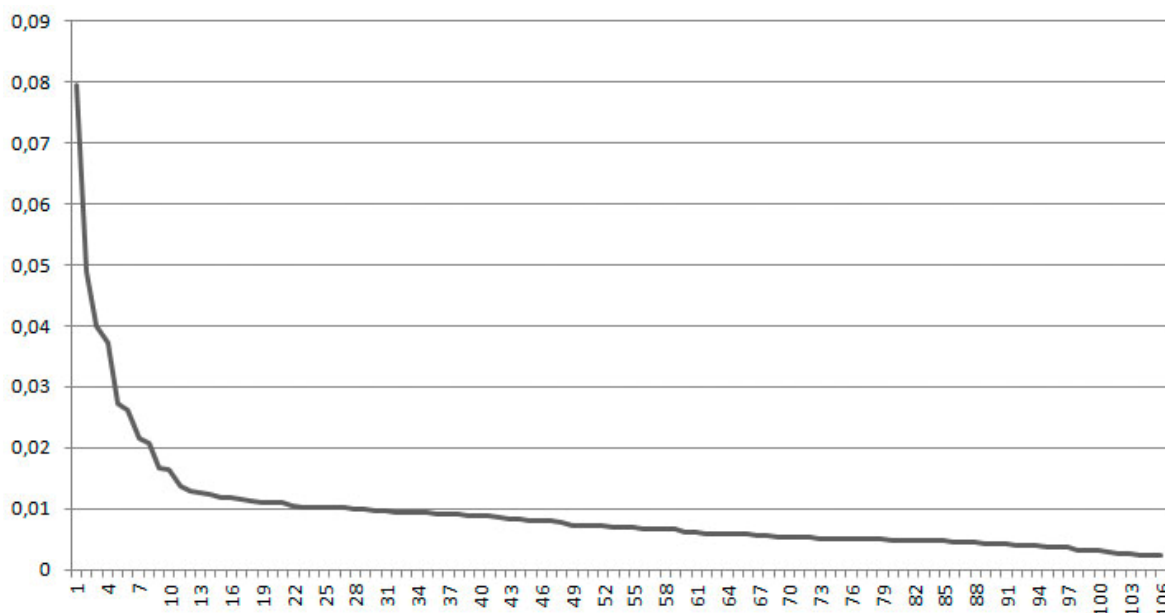


Рис. 1. Ранги вершин графа G_{106}

В настоящей работе моделируются несколько сценариев изменения исходного графа G_{106} и для каждого из них приводится сравнение рангов вершин в новых графах (в зависимости от типа новых ссылок). Кроме этого, находятся значения структурных параметров графов, при которых ранг новых вершин становится выше, чем ранг лидеров исходного графа. Приводятся базовые сведения о методе вычисления рангов вершин и рассматриваются сценарии, основанные на соглашениях с лидерами и аутсайдерами.

Алгоритм PageRank

Пусть $G = (V, E)$ – ориентированный граф с множеством вершин $V, |V| = n$, и множеством дуг E . Для $i = 1, 2, \dots, n$ обозначим через q_i число дуг, выходящих из вершины i . Определим $(n \times n)$ -матрицу $H = (h_{ij})$ следующим образом. Если $q_i \neq 0$, то положим для $j = 1, \dots, n$

$$h_{ij} = \begin{cases} 1/q_i, & \text{если имеется дуга из вершины } i \text{ в вершину } j; \\ 0, & \text{иначе.} \end{cases}$$

Если $q_i = 0$, то для $j = 1, \dots, n$ положим $h_{ij} = 1/n$. По определению матрицы H сумма элементов в каждой ее строке равна единице, т. е. H является стохастической матрицей. Рассмотрим $(n \times n)$ -матрицу E , каждый элемент которой равен $1/n$, и определим матрицу

$$S = \alpha H + (1 - \alpha)E,$$

где параметр α называется *коэффициентом затухания*. Обычно его значение принимают равным 0,85. Очевидно, все элементы матрицы S положительны и она является стохастиче-

ской. Матрица $S = (s_{ij})$ известна в литературе как *Google matrix*. Рангом вершины i назовем величину π_j , удовлетворяющую соотношению

$$\pi_j = \sum_{i=1}^n \pi_i s_{ij},$$

которое можно записать в виде $\pi = \pi \cdot S$. Вектор $\pi = (\pi_1, \dots, \pi_n)$ назовем *вектором рангов вершин* графа. Обычно вектор π нормализуется так, что $\sum_{i=1}^n \pi_i = 1$. В этом случае получаем уравнение

$$\pi = \alpha \pi S + (1 - \alpha)u, \quad \pi = \alpha \pi S + (1 - \alpha)u,$$

где n -вектор u имеет вид $(1/n, \dots, 1/n)$.

Соглашение с лидерами

Графы $[G_{106+n}]^k$ и $[G_{106+n}^*]^k$. Рассмотрим сценарии присоединения новых вершин к графу G_{106} , вершинами которого являются v_1, \dots, v_{106} , исходя из предположения, что лидеры графа G_{106} устанавливают ссылки на новые вершины.

Пусть число присоединяемых вершин равно $n, n \geq 1$. Будем предполагать, что новые вершины занумерованы u_1, \dots, u_n . Обозначим через K_n ориентированный граф на n вершинах u_1, \dots, u_n , в котором каждая вершина соединена ориентированной дугой с каждой другой вершиной. Таким образом, K_n имеет $n(n-1)$ дуг. Обозначим через G_{106+n} граф с множеством вершин $\{v_1, \dots, v_{106}\} \cup \{u_1, \dots, u_n\}$, множество дуг которого состоит из дуг графа G_{106} и дуг графа K_n .

Проведем ранжирование вершин графа G_{106+n} методом PageRank. Это ранжирование индуцирует ранжирование вершин на его подграфе G_{106} . Перенумеруем вершины подграфа G_{106} в соответствии с этим ранжированием в порядке невозрастания рангов: v_1, \dots, v_{106} . Ранг вершины v_i будем здесь и далее обозначать $\pi_i, i = 1, \dots, 106$.

Пусть k – целый параметр, принимающий значения от 0 до 106. Обозначим через $[G_{106+n}]^k$ граф, полученный из графа G_{106+n} добавлением дуг из вершин v_1, \dots, v_k в каждую из вершин u_1, \dots, u_n . В частности, $[G_{106+n}]^0$ совпадает с G_{106+n} .

Через $[G_{106+n}^*]^k$ обозначим граф, полученный из $[G_{106+n}]^k$ добавлением дуг из вершин u_1, \dots, u_n в каждую из вершин v_1, \dots, v_{106} , при этом $[G_{106+n}^*]^0 \equiv G_{106+n}^*$.

Отметим, что каждый из графов $K_n, [G_{106+n}]^k, [G_{106+n}^*]^k$ обладает автоморфизмом порядка n , переводящим $u_i \rightarrow u_{i+1}, i = 1, \dots, n$, где индексы берутся по модулю n , и оставляющим вершины $v_j, j = 1, \dots, 106$, на месте. Следовательно, в фиксированном графе вершины u_1, \dots, u_n имеют один и тот же ранг, который мы будем обозначать π_* .

Добавление одной вершины. Пусть к графу G_{106} присоединяется одна вершина, т. е. $n = 1$. Получаемый граф G_{107} есть несвязное объединение графа G_{106} и изолированной вершины u_1 . Проведем ранжирование вершин графа G_{107} (который совпадает с графом $[G_{107}]^0$) с помощью алгоритма PageRank и перенумеруем вершины подграфа G_{106} в соответствии с невозрастанием рангов. Наибольший ранг $\pi_1 = 0,079367825$ имеет вершина v_1 , соответствующая сайту Портала СО РАН, а ранг новой вершины u_1 в графе G_{107} равен $\pi_* = 0,002341948$ (табл. 1).

Таблица 1

Сценарии соглашений с лидерами

№	Граф		π_1	π_*	
1	$[G_{107}]^k$	$k = 0$	0,079367825	0,002341948	
		$k = 14$	0,076850950	0,015046168	π_* вошел в TOP10 рангов
		$k = 56$	0,070007676	0,070268180	u_1 стала лидером
		$k = 106$	0,052438192	0,179566009	
2	$[G_{107}^*]^k$	$k = 0$	0,079369311	0,002323365	
		$k = 14$	0,075984516	0,017483016	π_* вошел в TOP10 рангов
		$k = 56$	0,067356153	0,072655435	u_1 стала лидером
		$k = 106$	0,052512834	0,178390171	
3	$[G_{108}]^k$	$k = 0$	0,077139781	0,015174368	
		$k = 1$	0,076473629	0,019415946	π_* вошел в TOP10 рангов
		$k = 11$	0,068414878	0,068819123	u_1 и u_2 стали лидерами
		$k = 106$	0,018439751	0,332701182	
4	$[G_{108}^*]^k$	$k = 0$	0,079185308	0,002318151	
		$k = 19$	0,076473629	0,019415946	π_* вошел в TOP10 рангов
		$k = 62$	0,058551613	0,060915392	u_1 и u_2 стали лидерами
		$k = 106$	0,042524987	0,114182289	
5	$[G_{111}]^k$	$k = 0$	0,073780996	0,014513776	
		$k = 1$	0,072277316	0,019415946	π_* вошел в TOP10 рангов
		$k = 13$	0,055078863	0,056041798	u_1, \dots, u_5 стали лидерами
		$k = 106$	0,010322062	0,149987794	
6	$[G_{111}^*]^k$	$k = 0$	0,078638226	0,002302630	
		$k = 19$	0,067449214	0,015512883	π_* вошел в TOP10 рангов
		$k = 94$	0,044663227	0,045460593	u_1, \dots, u_5 стали лидерами
		$k = 106$	0,028538459	0,061726068	
7	$[G_{131}]^k$	$k = 0$	0,057182244	0,011248646	
		$k = 1$	0,052941540	0,013411782	π_* вошел в TOP10 рангов
		$k = 16$	0,026390432	0,025116797	u_1, \dots, u_{25} стали лидерами
		$k = 106$	0,003111032	0,034032701	
8	$[G_{131}^*]^k$	$k = 0$	0,075171288	0,002203713	
		$k = 38$	0,011272822	0,015512883	π_* вошел в TOP10 рангов
		$k = 79$	0,015871734	0,016937550	u_1, \dots, u_{25} стали лидерами
		$k = 106$	0,011272822	0,018377486	

Сценарий 1. Естественно предположить, что ранг вершины u_1 будет увеличиваться по мере того, как на нее будет ссылаться все большее число лидеров. Предположим, что на u_1 сделали ссылки первые k лидеров подграфа G_{106} , а именно вершины v_1, \dots, v_k . Этот сценарий моделируется графом $[G_{107}]^k$. Проведя вычисления рангов вершин графов $[G_{107}]^k$ для $k = 1, \dots, 106$, заключаем, что ранг вершины u_1 войдет в десятку наибольших рангов при $k = 14$ и u_1 станет лидером графа $[G_{107}]^k$ при $k = 54$ (см. табл. 1, где для каждого фиксированного k ранг вершины v_1 обозначен через π_1 , а ранг вершины u_1 – через π_*).

Сценарий 2. Предположим, что мы не ограничиваемся соглашениями с k лидерами, а, управляя сайтом, соответствующим вершине u_1 , устанавливаем ссылки из вершины u_1 на все вершины подграфа G_{106} , т. е. проводим из вершины u_1 дуги во все остальные вершины. Этот сценарий моделируется графами $[G_{107}^*]^k$, $k = 1, \dots, 106$. Вычисляя ранги вершин этих графов, получаем, что ранг вершины u_1 войдет в десятку наибольших рангов при $k = 14$, и u_1 станет лидером графа $[G_{107}^*]^k$ при $k = 56$ (см. табл. 1).

Коалиция двух вершин. Как видно из моделирования сценариев 1 и 2, для достижения лидерских позиций вершиной u_1 требуется добиться соглашений о ссылках со всеми сайтами из первой половины списка, составленного по рангам. Покажем, что применение других сценариев может привести к нужному результату с меньшими затратами, а именно: рассмотрим сценарии создания коалиции с еще одной вершиной, стремящейся к попаданию в лидеры. Предположим, что к графу G_{106} добавлены две новые вершины u_1 и u_2 , имеющие ссылки друг на друга. Полученный граф G_{108} соответствует описанной выше конструкции при $n = 2$. Ранжирование его вершин с помощью алгоритма PageRank показывает, что наибольший ранг $\pi_1 = 0,077139781$ имеет вершина, соответствующая сайту Портала СО РАН, а ранги новых вершин u_1 и u_2 совпадают и равны $\pi_* = 0,015174368$ (см. табл. 1).

Сценарий 3. Предположим, что на вершины u_1 и u_2 сделали ссылки первые k лидеров подграфа G_{106} графа G_{108} (обозначенные после ранжирования вершин графа G_{108} как v_1, \dots, v_k). Этот сценарий моделируется графом $[G_{108}]^k$. Вычисляя ранги вершин графов $[G_{108}]^k$, где $k = 1, \dots, 106$, получаем, что ранг π_* вершин u_1 и u_2 войдет в десятку самых высоких рангов при $k = 1$ и вершины u_1, u_2 станут лидерами графа $[G_{108}]^k$ при $k = 11$ (см. табл. 1). Таким образом, согласованность действий двух новых вершин существенно повышает ранг каждой из них.

Сценарий 4. Предположим, что мы не ограничиваемся соглашениями с k лидерами, а, управляя сайтами, соответствующими вершинам u_1 и u_2 , устанавливаем с них ссылки на все вершины веб-графа G_{106} , т. е. из вершин u_1 и u_2 проводим дуги во все остальные вершины. Этот сценарий моделируется графами $[G_{108}^*]^k$, $k = 1, \dots, 106$. Вычисляя ранги вершин этих графов, получаем, что ранг π_* вершин u_1 и u_2 войдет в первую десятку рангов при $k = 19$ и эти вершины станут лидерами графа $[G_{108}^*]^k$ при $k = 62$ (см. табл. 1).

Коалиция большой группы вершин. Рассмотрим сценарии создания коалиций с большим числом новых вершин, ставящими перед собой задачу выйти в лидеры, исходя из предположения, что стратегии всех новых вершин одинаковы.

Сценарий 5. Предположим, что к графу G_{106} присоединяются пять новых вершин u_1, \dots, u_5 , имеющие ссылки друг на друга. Результатом такого присоединения является граф G_{111} , соответствующий описанной выше конструкции при $m = 5$. Проведем ранжирование его вершин с помощью алгоритма PageRank. Наибольший ранг $\pi_1 = 0,073780996$ имеет вершина v_1 , соответствующая сайту Портала СО РАН, а ранги новых вершин u_1, \dots, u_5 совпадают и равны $\pi_* = 0,014513776$ (см. табл. 1). Предположим, что на вершины u_1, \dots, u_5 сделали ссылки первые k лидеров подграфа G_{106} графа G_{111} . Этот сценарий моделируется графом $[G_{111}]^k$. Вычисляя ранги для вершин графов $[G_{111}]^k$, где $k = 1, \dots, 106$, получаем, что ранг π_* вершин u_1, \dots, u_5 войдет в десятку самых высоких рангов при $k = 1$ и вершины u_1, \dots, u_5 станут лидерами

ми графа $[G_{111}]^k$ при $k = 13$ (см. табл. 1). Таким образом, согласованность действий пяти новых вершин существенно повышает ранг каждой из них.

Проиллюстрируем сценарий 5 сравнением графиков рангов вершин графов G_{111} и $[G_{111}]^{20}$. На рис. 2 приведены графики ранжирования вершин графа G_{111} и графа $[G_{111}]^{20}$, соответствующего случаю, когда на каждую из вершин u_1, \dots, u_5 ссылаются 20 лидеров графа G_{111} . Упорядочивание вершин соответствует их рангам в графе G_{111} . Как видно, вершины u_1, \dots, u_5 оказались в нем на местах с 11 по 15. В графе $[G_{111}]^{20}$ имеем $\pi_1 = 0,045291$ и ранг вершин u_1, \dots, u_5 равен $\pi_* = 0,074831$. Новые вершины стали лидерами с большим отрывом.

Сценарий 6. Рассмотрим сценарий присоединения пяти новых вершин, который отличается от предыдущего сценария тем, что из каждой новой вершины устанавливаются ссылки на старые вершины графа. Этот сценарий моделируется графами $[G_{111}^*]^k$, $k = 1, \dots, 106$. Вычисляя ранги вершин этих графов, получаем, что ранги вершин u_1, \dots, u_5 войдут в первую десятку рангов при $k = 19$ и эти вершины станут лидерами графа $[G_{111}^*]^k$ при $k = 94$ (см. табл. 1).

Проиллюстрируем сценарий 6 сравнением графиков рангов вершин графов. На рис. 3 приведены графики рангов вершин графов G_{111} и G_{111}^* . Вершины упорядочены в соответствии с их рангами в графе G_{111} . На рис. 4 приведены графики рангов вершин графов $[G_{111}]^{20}$ и $[G_{111}^*]^{20}$. Здесь вершины упорядочены в соответствии с их рангами в графе $[G_{111}]^{20}$.

Сценарий 7. Предположим, что к графу G_{106} присоединяется достаточно большое число вершин. Приведем результаты вычислений в случае присоединения двадцати пяти вершин (почти четверть от числа старых вершин). Положим, что действия новых вершин u_1, \dots, u_{25} согласованы и каждая из них ссылается на другую. Результатом такого присоединения является граф G_{131} , соответствующий описанной выше конструкции при $n = 25$. Проведем ранжирование его вершин с помощью алгоритма PageRank. Наибольший ранг $\pi_1 = 0,057182244$ имеет вершина v_1 , соответствующая сайту Портала СО РАН, а ранги новых вершин u_1, \dots, u_{25} совпадают и равны $\pi_* = 0,011248646$ (см. табл. 1). Предположим, что на вершины

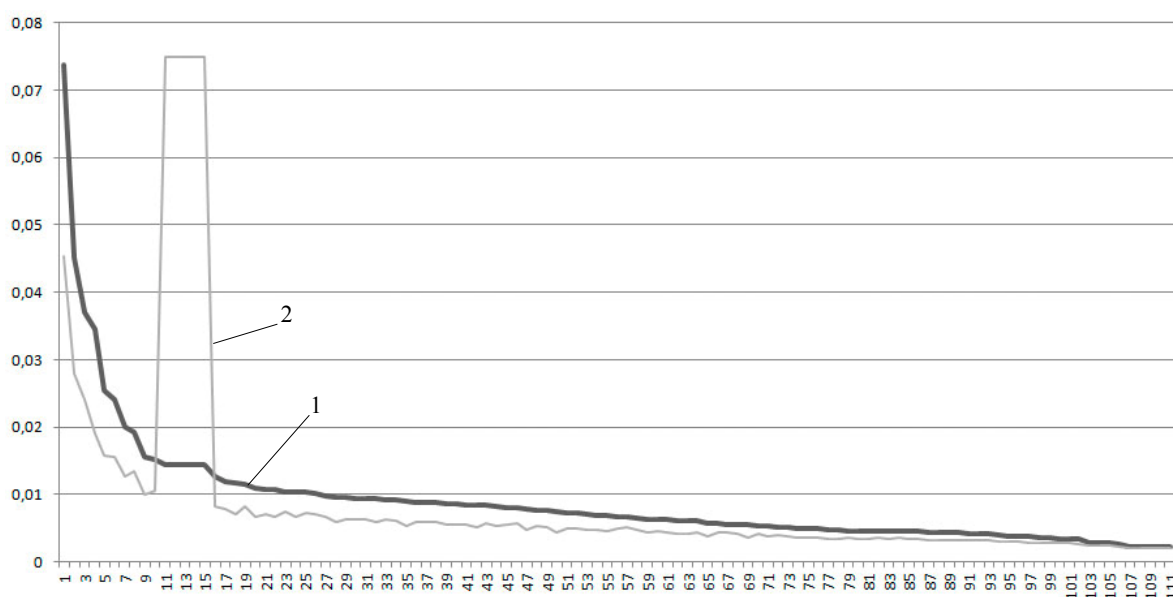


Рис. 2. Ранги вершин графа G_{111} (1) и графа $[G_{111}]^{20}$ (2)

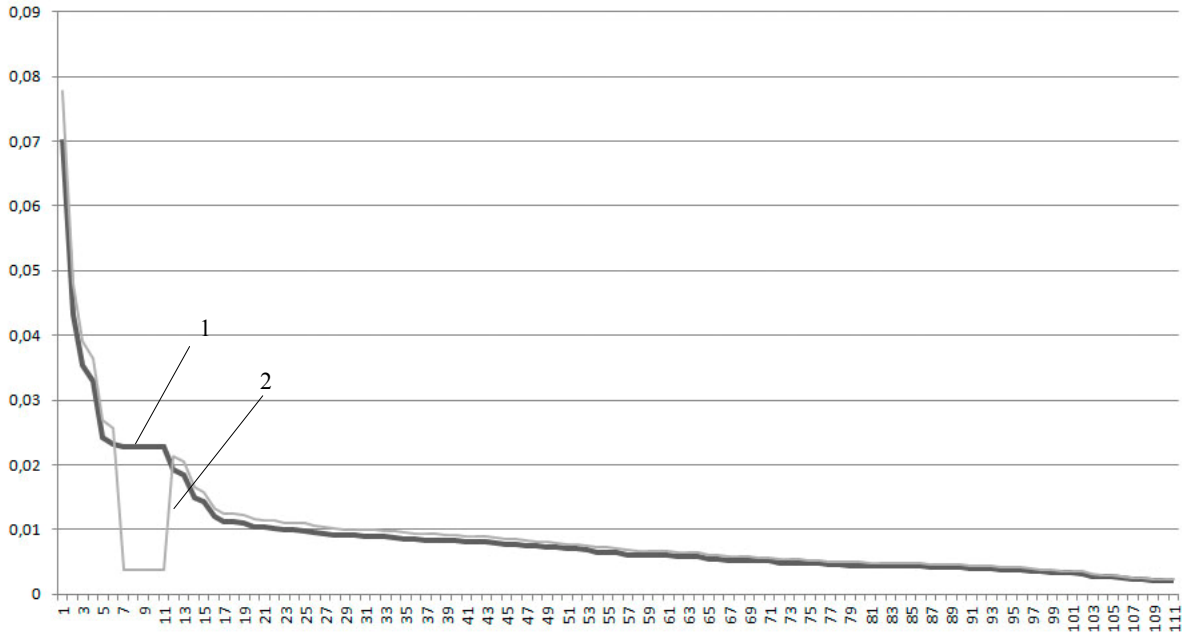


Рис. 3. Ранги вершин графа G_{111} (1) и графа G_{111}^* (2)

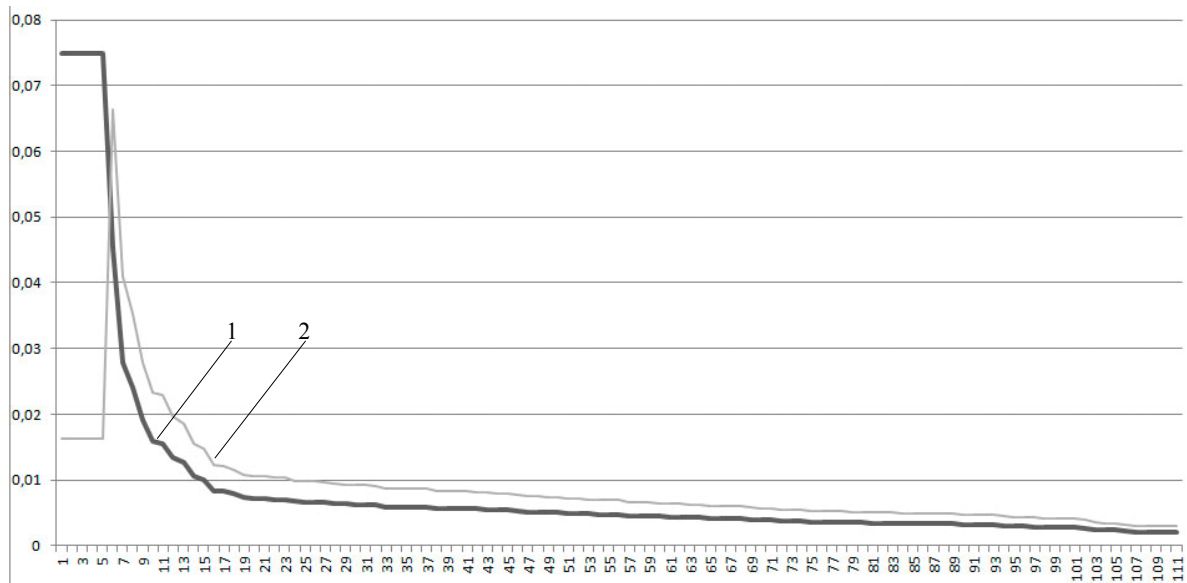


Рис. 4. Ранги вершин графа $[G_{111}]^{20}$ (1) и графа $[G_{111}^*]^{20}$ (2)

u_1, \dots, u_{25} сделали ссылки первые k лидеров подграфа G_{106} графа G_{131} . Этот сценарий моделируется графом $[G_{131}]^k$. Вычисляя ранги вершин графов $[G_{131}]^k$, где $k = 1, \dots, 106$, получаем, что ранг π_* вершин u_1, \dots, u_{25} войдет в десятку самых высоких рангов при $k = 1$, а вершины u_1, \dots, u_{25} станут лидерами графа $[G_{131}]^k$ при $k = 16$ (см. табл. 1).

Сценарий 8. Рассмотрим сценарий присоединения двадцати пяти новых вершин, который отличается от предыдущего сценария тем, что из каждой новой вершины устанавливаются ссылки на старые вершины графа. Этот сценарий моделируется графами $[G_{131}^*]^k$,

$k = 1, \dots, 106$. Вычисляя ранги вершин этих графов, получаем, что ранг π_* вершин u_1, \dots, u_{25} войдет в первую десятку рангов при $k = 38$ и эти вершины станут лидерами графа $[G_{131}^*]^k$ при $k = 79$ (см. табл. 1).

Соглашение с аутсайдерами

Рассмотрим сценарии присоединения новых вершин к графу G_{106} исходя из предположения, что ссылки на новые вершины графа G_{106} устанавливают аутсайдеры. Такие сценарии связаны, например, с ситуациями, когда лидеров трудно убедить в необходимости установления ссылок на новые вершины или когда решение вопроса об установлении ссылок от лидеров требует слишком больших затрат. Естественно предполагать, что с аутсайдерами рейтинга решить вопрос об установлении ссылок на новые вершины может оказаться проще и / или дешевле.

Как и ранее, будем присоединять $n \geq 1$ вершин u_1, \dots, u_n , образующих ориентированный граф K_n , в котором каждая вершина соединена ориентированной дугой с каждой другой вершиной. Обозначим через G_{106+n} граф, вершинами которого являются v_1, \dots, v_{106} вместе с u_1, \dots, u_n , а множество дуг состоит из дуг графа G_{106} и дуг графа K_n .

Проведем ранжирование вершин графа G_{106+n} методом PageRank. Это ранжирование индуцирует ранжирование вершин на его подграфе G_{106} . Перенумеруем вершины подграфа G_{106} в соответствии с этим ранжированием в порядке невозрастания рангов: v_1, \dots, v_{106} , т. е. вершина v_1 имеет наибольший ранг (является лидером), а вершина v_{106} – наименьший (является аутсайдером). Пусть ℓ – целый параметр, принимающий значения от 0 до 106, и пусть граф $[G_{106+n}]_\ell$ получен из графа G_{106+n} добавлением дуг из ℓ вершин $v_{106-(\ell-1)}, \dots, v_{106}$ в каждую из вершин u_1, \dots, u_n . В частности, $[G_{106+n}]_0$ совпадает с G_{106+n} . Через $[G_{106+n}^*]_\ell$ обозначим граф, полученный из $[G_{106+n}]_\ell$ добавлением дуг из вершин u_1, \dots, u_n в каждую из вершин v_1, \dots, v_{106} . В частности, граф $[G_{106+n}^*]_0$ совпадает с графом G_{106+n}^* .

Сценарии 9–16. Рассмотрим сценарии 9–16, которые отличаются от описанных выше сценариев 1–8 лишь тем, что дуги в новые вершины проводятся не из k лидеров, а из ℓ аутсайдеров. Следовательно, эти сценарии моделируются не графами $[G_{106+n}]^k$ и $[G_{106+n}^*]^k$, а графами $[G_{106+n}]_\ell$ и $[G_{106+n}^*]_\ell$. Результаты вычислений рангов вершин этих графов приведены в табл. 2, где π_1 – ранг лидера подграфа G_{106} , а π_* – ранг новых вершин u_1, \dots, u_n в графах $[G_{106+n}]_\ell$ и $[G_{106+n}^*]_\ell$.

Таблица 2

Сценарии соглашений с аутсайдерами

№	Граф		π_1	π_*	
9	$[G_{107}]_\ell$	$\ell = 0$	0,079367825	0,002341948	
		$\ell = 8$	0,077964999	0,018282575	π_* вошел в TOP10 рангов
		$\ell = 32$	0,069065081	0,069436005	u_1 стала лидером
		$\ell = 106$	0,052438192	0,179566009	
10	$[G_{107}^*]_\ell$	$\ell = 0$	0,079369311	0,002323365	
		$\ell = 8$	0,077976340	0,018139774	π_* вошел в TOP10 рангов
		$\ell = 32$	0,068832481	0,072569971	u_1 стала лидером
		$\ell = 106$	0,052512834	0,178390171	

Окончание табл. 2

№	Граф		π_1	π_*	
11	$[G_{108}]_\ell$	$\ell = 0$	0,077139781	0,015174368	
		$\ell = 1$	0,076521508	0,017727548	π_* вошел в TOP10 рангов
		$\ell = 10$	0,068414878	0,068819123	u_1 и u_2 стали лидерами
		$\ell = 106$	0,018439751	0,332701182	
12	$[G_{108}^*]_\ell$	$\ell = 0$	0,079185308	0,002318151	
		$\ell = 15$	0,075870437	0,018248326	π_* вошел в TOP10 рангов
		$\ell = 47$	0,060898092	0,061400769	u_1 и u_2 стали лидерами
		$\ell = 106$	0,042524987	0,114182289	
13	$[G_{111}]_\ell$	$\ell = 0$	0,073780996	0,014513776	
		$\ell = 1$	0,072865177	0,016020083	π_* вошел в TOP10 рангов
		$\ell = 18$	0,054662124	0,055686559	u_1, \dots, u_5 стали лидерами
		$\ell = 106$	0,010322062	0,149987794	
14	$[G_{111}^*]_\ell$	$\ell = 0$	0,078638226	0,00230263	
		$\ell = 26$	0,064710110	0,015532979	π_* вошел в TOP10 рангов
		$\ell = 68$	0,043012516	0,043275725	u_1, \dots, u_5 стали лидерами
		$\ell = 106$	0,028538459	0,061726068	
15	$[G_{131}]_\ell$	$\ell = 0$	0,057182244	0,011248646	
		$\ell = 2$	0,055678455	0,011830183	π_* вошел в TOP10 рангов
		$\ell = 40$	0,024277944	0,024334259	u_1, \dots, u_{25} стали лидерами
		$\ell = 106$	0,003111032	0,034032701	
16	$[G_{131}^*]_\ell$	$\ell = 0$	0,075171288	0,002203713	
		$\ell = 62$	0,033145778	0,011273166	π_* вошел в TOP10 рангов
		$\ell = 91$	0,016934049	0,016938960	u_1, \dots, u_{25} стали лидерами
		$\ell = 106$	0,011272822	0,018377486	

Выводы

В работе рассмотрены сценарии присоединения новых вершин к веб-графу научных организаций Сибирского отделения Российской академии наук. Сценарии подразумевали как присоединение одной вершины, так и присоединение коалиции вершин, действующих согласованно. Первая группа сценариев была связана с установлением ссылок на новые вершины от лидеров исходного графа, а вторая – с установлением ссылок от аутсайдеров исходного графа. В результате построения моделей для рассматриваемых сценариев и ранжирования вершин графов методом PageRank установлено, что сценарий согласованного действия двух вершин при соглашениях с аутсайдерами является оптимальным: новые вершины становятся лидерами графа после установления на них всего лишь десяти ссылок от аутсайдеров исходного графа.

Список литературы

1. Шокин Ю. И., Клименко О. А., Петров И. С. Анализ связей между институтами Сибирского отделения РАН // Вестн. Новосиб. гос. ун-та. Серия: Информационные технологии. 2011. Т. 9, вып. 4. С. 12–17.
2. Шокин Ю. И., Веснин А. Ю., Добрынин А. А., Клименко О. А., Рычкова Е. В., Петров И. С. Исследование научного веб-пространства Сибирского отделения Российской академии наук // Журнал вычислительных технологий. 2012. Т. 17, №. 6. С. 85–98.

3. Brin S., Page L. The Anatomy of a Large-Scale Hypertextual Web Search Engine // Comput. Netw. ISDN Syst. 1998. Vol. 30. No. 1–7. P. 107–117.
4. Ding Y. Topic-Based PageRank on Author Co-citation Networks // Journal of the American Society for Information Science and Technology. 2011. Vol. 62 (3). P. 449–466.
5. Pandurangan G., Raghavan P., Upfal E. Using PageRank to Characterize Web Structure // Internet Mathematics. 2006. Vol. 3. No. 1. P. 1–20.
6. Константинова Е. В., Савин М. Ю., Клименко О. А. Анализ научного веб-пространства СО РАН методом PageRank // Распределенные информационные и вычислительные ресурсы: Материалы XIV Рос. конф. с участием иностр. ученых. Новосибирск, 2012. URL: <http://conf.nsc.ru/dicr2012/ru/reportview/141956>
7. Константинова Е. В., Савин М. Ю., Клименко О. А. Анализ динамики развития Веб-пространства Сибирского отделения РАН методом PageRank // Информационные и математические технологии в науке и управлении // Информационные и математические технологии в науке и управлении: Тр. XVIII Байкальской Всерос. конф. Иркутск, 2013. Т. 3. С. 142–148.

Материал поступил в редколлегию 07.10.2013

A. Yu. Vesnin, E. V. Konstantinova, M. Yu. Savin

ON SCENARIOS OF JOINING NEW SITES TO THE SB RAS WEB-SPACE

We consider transformations of the Web-graph of the Siberian Branch of Russian Academy of Sciences (SB RAS). The vertices in this graph correspond to the web-pages of scientific institutes, and there is a directed edge between two vertices if one of the web-pages is linking to another web-page. Different scenarios for joining new vertices to this Web-graph are considered and changes in ranging of vertices by PageRank are investigated. Structural parameters of graphs making a new vertex to be the leader are shown.

Keywords: Web-graph, Web-Space, PageRank, vertex rank, site.