ФУНДАМЕНТАЛЬНЫЕ МАТЕМАТИЧЕСКИЕ ОСНОВЫ ТЕОРИИ УПРАВЛЕНИЯ

УДК 517.977
ББК 22.176

ВЛИЯНИЕ В СЕТЕВЫХ СТРУКТУРАХ С ИСПОЛЬЗОВАНИЕМ ИНДЕКСОВ ДАЛЬНИХ ВЗАИМОДЕЙСТВИЙ

Алескеров Ф. Т.[1], Мещерякова Н. Г.[2], Швыдун С. В.[3]

(*Национальный исследовательский университет «Высшая школа экономики», Институт проблем управления РАН, Москва*)

Рассматривается задача оценки влияния в сетевых структурах. Предложено несколько моделей, позволяющих учитывать индивидуальные характеристики каждой вершины, а также интенсивность ближних и дальних взаимодействий между ними. Интенсивность взаимодействий оценивается как с помощью анализа всех возможных каналов влияния одной вершины на другую, так и с помощью подхода, в основе которого лежат симуляции. Предложенные модели помогает выявить как очевидные, так и скрытые центральные элементы, которые не могу быть определены классическими

[1] *Фуад Тагиевич Алескеров, доктор технических наук, ординарный профессор, заведующий МЛАВР НИУ ВШЭ, руководитель департамента математики факультета экономических наук НИУ ВШЭ, заведующий лабораторией №25 ИПУ РАН* (alesk@hse.ru).

[2] *Наталья Геннадьевна Мещерякова, студент, стажер-исследователь МЛАВР НИУ ВШЭ, техник лаборатории №25 ИПУ РАН (natamesc@gmail.com).*

[3] *Сергей Владимирович Швыдун, магистр бизнес-информатики, стажер-исследователь МЛАВР НИУ ВШЭ, преподаватель департамента математики факультета экономических наук НИУ ВШЭ, младший научный сотрудник лаборатории №25 ИПУ РАН (shvydun@hse.ru).*

мерами центральностей или другими индексами оценки влияния.

Ключевые слова: анализ сетей, индексы влияния, ближние и дальние взаимодействия.

1. Введение

В последнее время все больший интерес уделяется анализу различных сообществ и сложных сетей, в частности исследованию их структуры и определению ключевых элементов. На данный момент существует большое количество индексов центральности, с помощью которых можно оценить степень влияния каждой вершины сети. Широко известны степенные центральности, учитывающие число связей и их интенсивность для каждой вершины [5], центральности по близости и посредничеству, в основе которых лежит поиск кратчайших путей между элементами сети [6,7,9,10], а также собственно-векторные центральности, которые оценивают важность вершины с учетом важности ее соседей [3,4,8]. К сожалению, большинство из них не позволяют учитывать важные характеристики – индивидуальные атрибуты каждой вершины. Более того, существующие методы не в полной мере оценивают интенсивность дальних взаимодействий и зачастую учитывают незначимые связи между элементами сети. Следовательно, результаты применения классических индексов центральности не в полной мере отображают реальную картину, так как с их помощью недооценить действительно важных элементов сети.

В работе [2] был предложен новый метод оценки влияния в сетях, в основе которого лежит расчета индекса, разработанного в [1]. Отличительной его особенностью является то, что он учитывает такие индивидуальные характеристики вершины, как ее размер (силу) и критическое пороговое значение, при достижении которого другие вершины начинают на нее влиять. Тем не менее, разработанный индекс учитывает исключительно ближние взаимодействия между вершинами, а степень влияния вершин, не имеющих друг с другом прямых связей, оставалось нулевой.

В нашей работе предложен ряд моделей, позволяющих учесть существующие недостатки в работе [2]. Отличительной особенностью наших моделей является то, что они позволяют оценить влияние вершин, не имеющих между собой прямых связей и находящихся на расстоянии *s*, значение которого может варьироваться от задачи. Интенсивность дальних взаимодействий может быть оценена как с помощью анализа всех возможных каналов влияния одной вершины на другую, так и с помощью подхода, в основе которого лежат симуляции.

2. Индексы дальних взаимодействий

2.1 ПОСТАНОВКА ЗАДАЧИ

Рассмотрим граф $G=\{V,E,W\}$, где $V=\{1,...,n\}$ - множество вершин, $|V|=N$, $E\subseteq V\times V$ - множество ребер, а $W=\{|w_{ij}|\}$ - множество весов - действительных чисел, соответствующих каждому ребру $(i,j)\in E$. Граф *G* называется направленным, если $\forall i,j\in V:(i,j)\in E\Rightarrow(j,i)\notin E$ и ненаправленным в противном случае. Граф *G* называется взвешенным, если $\forall i_1,i_2,j_1,j_2\in V:(i_1,j_1)\in E$ & $(i_2,j_2)\in E$ $\Rightarrow w_{i_1 j_1}=w_{i_2 j_2}$. В рамках нашей работы мы рассматриваем только направленные взвешенные графы.

Граф *G* может быть также представлен с помощью матрицы смежности $A=[a_{ij}]_{N\times N}$, где $a_{ij}=1$, если $(i,j)\in E$ и $a_{ij}=0$ в ином случае, а также с помощью матрицы весов $W=[w_{ij}]_{N\times N}$, где w_{ij} отражает вес ребра $(i,j)\in E$. В терминах влияния, $(i,j)\in E$ показывает наличие прямого влияния вершины *i* на вершину *j*, а w_{ij} - интенсивность влияния. Дополнительно каждая вершина может иметь набор атрибутов w_i^k, где *i* – номер вершины, *k* – номер атрибута, $k\in K$, а также пороговое значение q_i, при достижении которого вершина i подвержена влиянию других вершин.

Обозначим также через N_i набор вершин, имеющих прямую связь с вершиной i. Тогда группа вершин $\Omega(i) \subseteq N_i$ является критической, если $\sum_{j \in \Omega(i)} w_{ji} \geq q_i$, а вершина $l \in \Omega(i)$ - ключевой, если ее исключение из данной группы делает группу некритической. Набор ключевых вершин в группе $\Omega(i)$ обозначим через $\Omega_p(i)$. Соответственно, необходимо оценить степень влияния каждой вершины на другие вершины сети.

2.2. ИНДЕКСЫ ДАЛЬНИХ ВЗАИМОДЕЙСТВИЙ, ОСНОВАННЫХ НА ПУТЯХ МЕЖДУ ВЕРШИНАМИ

Построим матрицу $C = [c_{ij}]_{N \times N}$ на основе матрицы W и заранее определенной величины порога, показывающую наихудший сценарий влияния прямого влияния одной вершины на другую

$$c_{ij} = \begin{cases} \dfrac{w_{ij}}{\min\limits_{\Omega(j) | i \in \Omega_p(j)} \sum_{l \in \Omega(j)} w_{lj}}, i \in \Omega_p(i) \\ 0, i \notin \Omega_p(i). \end{cases}$$

Матрица C может быть интерпретирована следующим образом. Если $c_{ij} = 1$, тогда вершина i оказывает максимальное влияние на j. Наоборот, если $c_{ij} = 0$, то вершина i не оказывает влияния на вершину j. Если же $0 < c_{ij} < 1$, то мы рассматриваем случай максимального в терминах групп влияния вершины i на вершину j.

Таким образом, мы оцениваем прямое влияние элементов системы друг на друга. Для определения косвенного влияния двух элементов необходимо рассмотрим все возможные пути между ними. Напомним, что путем из i в j называется такая упорядоченная последовательность элементов i, j_1, …, j_k, j, что $i\rho j_1$, $j_1\rho j_2$, …, $j_{k-1}\rho j_k$, $j_k\rho j$, где $j_1 \rho j_2 \Leftrightarrow c_{j_1 j_2} > 0$.

Рассмотрим все возможные пути между вершинами, не содержащие циклов. Обозначим через $P^{ij} = \{P_1^{ij}, P_2^{ij}, ..., P_m^{ij}\}$ набор всех возможных путей между вершинами i и j. Тогда непрямое влияние вершины i на вершину j через путь P_k^{ij} можно оценить следующим образом

(1) $f(P_k^{ij}) = c_{ij(1,k)} \cdot c_{j(1,k)j(2,k)} \cdot ... \cdot c_{j(n(k),k)j}$

или

(2) $f(P_k^{ij}) = \min(c_{ij(1,k)}, c_{j(1,k)j(2,k)}, ..., c_{j(n(k),k)j})$,

где $n(k)$ - длина пути P_k^{ij}, $j(l,k)$ - номер вершины, которая находится на l-ой позиции пути P_k^{ij} .

Формулы (1) и (2) могут быть интерпретированы следующим образом. Согласно формуле (1) совокупное влияние элемента i на элемент j рассчитывается с учетом всех прямых влияний между элементами на этом пути, тогда как формула (2) определяет совокупное влияние как значение минимального прямого влияния на этом пути.

Так как между двумя вершинами может быть большое число путей, существует проблема их агрегирования для итоговой оценки влияния вершины i на вершину j. В нашей работе предложено несколько способов их агрегирования, с помощью которых можно сформировать матрицу влияния $C^*(s) = [c_{ij}^*(s)]_{N \times N}$, которая зависит от параметра s (ограничение на длину путей*)*

(3) $c_{ij}^*(s) = \min(1, \sum_{k:n(k) \le s} f(P_k^{ij}))$

(4) $c_{ij}^*(s) = \max_{k:n(k) \le s} f(P_k^{ij})$

Таким образом, комбинируя формулы (1) и (2) с формулами (3) и (4), для каждого элемента мы можем определить его косвенное влияние на любой другой элемент. Полученные результаты могут быть агрегированы в один вектор, который будет отражать общее влияние каждого элемента в системе. Агрегирование индивидуальных влияний может быть осуществлено на основе относительных весов или по отношению к любым другим параметрам w_i^k.

2.2. ИНДЕКСЫ ДАЛЬНИХ ВЗАИМОДЕЙСТВИЙ, ОСНОВАННЫХ НА СИМУЛЯЦИЯХ

Другим способом оценки влияния вершин сети является метод, в основе которого лежит симуляционный подход. Допустим, некоторый набор вершин начинает оказывать влияние на другие вершины. Сформируют ли он критическую группу для какой-то из вершин и приведет ли этот факт в свою очередь к цепной реакции образования новых критических групп?

Построим матрицу $C = [c_{ij}]_{N\times N}$ на основе матрицы W и заранее определенной величины порога, показывающую наихудший сценарий влияния прямого влияния одной вершины на другую

$$c_{ij} = \begin{cases} 1, w_{ij} \geq q_j, \\ \dfrac{w_{ij}}{q_j}, 0 < w_{ij} < q_j, \\ 0, w_{ij} = 0. \end{cases}$$

Аналогично предыдущему, группа вершин $\Omega(j) \subseteq N_j$ является критической, если $\sum_{i\in\Omega(j)} c_{ij} \geq 1$. Если группа $\Omega(j)$ существует, то вершина *j* подвержена влиянию данной группы и, следовательно, присоединяется к ней и оказывает влияние на другие вершины.

Симуляционный подход основан на следующей идее. Рассмотрим группу вершин (общее число k_0) и проанализируем, для каких вершин (общее число k_1) данная группа является критической. Далее, проанализируем, для какого набора вершин (общее число k_2) группа вершин k_0+k_1 сформирует критическую группу. Данная процедура повторяется до тех пор, пока не пройдено заданное число шагов *s*, либо не будет такого шага *r*, что k_r=0. Таким образом, можно определить набор вершин (общее число $\sum_{l=1}^{\min(s,r)} k_l$), на который было оказано влияние группой k_0.

Аналогично, можно выбрать другую группу вершин k_0' и проанализировать, на какие вершины повлияла данная группа. Таким образом, можно сформировать матрицу влияния $C*(s) = [c_{ij}^*(s)]_{N\times N}$, которая показывает, в каком проценте случаев вершина, выбранная на начальном этапе, повлияла на другие вершины сети.

Полученные результаты могут быть агрегированы в один вектор аналогично предыдущему методу.

3. Пример

Рассмотрим пример на Рис. 1. Допустим пороговое значение для каждой вершины одинаково и равно $q_j = 25$. Стрелки на Рис. 1 показывают степень зависимости одной вершины от другой.

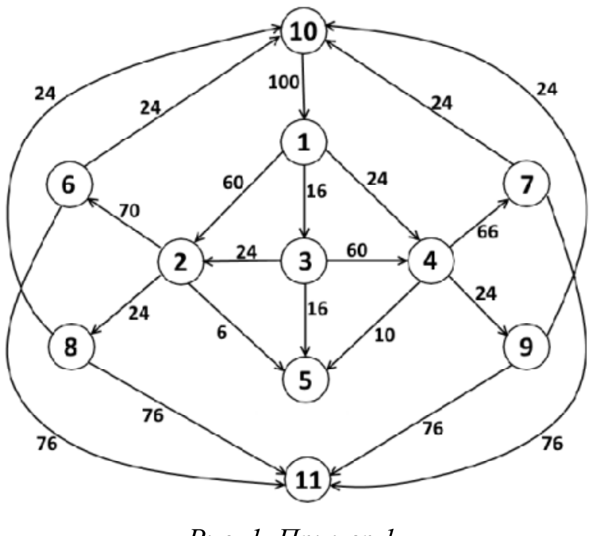

Рис. 1. Пример 1

Произведем расчет классических индексов центральности, индекса ближних взаимодействий (обозначим через SRIC), а также предложенных индексов (обозначим через LRIC). Полученные результаты приведем в Таблице 1.

Таблица 1. Индексы влияния для примера 1

№	**Степенная центральность**	**Центральность по близости**	**Центральность по посредничеству**	**PageRank**	**Собственно-векторная центральность**	**SRIC**	**LRIC (пути)**	**LRIC (симуляции)**
1	**200**	**0.0014**	**45**	0.61	**0.11**	**0.11**	0.03	0.03
2	184	**0.0012**	**23**	0.57	**0.1**	0.1	0.08	0.09
3	116	0.001	0	0.28	0.05	0.01	0.02	0
4	184	0.001	17	0.47	0.08	**0.11**	0.09	0.1
5	32	0.001	0	0.07	0.05	0.01	0.08	0.1
6	170	0.001	10	**0.7**	0.095	0.1	**0.11**	**0.12**
7	166	0.001	10	**0.65**	0.08	0.1	**0.12**	**0.14**
8	124	0.001	0	0.56	0.06	0.01	0.1	0.06
9	124	0.001	0	0.55	0.05	0.01	0.09	0.06
10	**196**	**0.0012**	**43**	0.64	0.09	0	0	0
11	**304**	**0.002**	0	**1**	**0.22**	**0.44**	**0.27**	**0.3**

Как показано выше, результаты предложенных моделей считают вершины №6,7,11 наиболее значимыми в сети. Наиболее похожие результаты показывает PageRank. Особое внимание необходимо уделить вершине №10. Согласно классическим индексам центральности, ее влияние достаточно высокое, однако, если внимательно проанализировать ситуацию, то оказывается, она данная вершина не оказывает влияния на другие вершины. Данная особенность выделяется предложенными индексами.

4. Заключение

В работе предложено несколько моделей, позволяющих учитывать индивидуальные характеристики каждой вершины, а также интенсивность ближних и дальних взаимодействий между ними. Продемонстрирован пример, показывающий отличие результатов предложенных моделей от результатов классических индексов центральности и других индексов.

Литература

1. Aleskerov F.T. Power indices taking into account agents' preferences In: B. Simeone & F. Pukelsheim (eds), Mathematics and Democracy, Berlin: Springer, pp. 1-18, 2006;
2. Aleskerov F.T., Andrievskaya I.K., Permjakova E.E. Key borrowers detected by the intensities of their short-range interactions / Working papers by NRU Higher School of Economics. Series FE "Financial Economics". 2014. No. WP BRP 33/FE/2014;
3. Bonacich P. Technique for Analyzing Overlapping Memberships// Sociological Methodology, Vol.4, 1972, pp.176-185;
4. Brin S., Page, L. The anatomy of a large-scale hypertextual Web search engine// Comput. Netw., 30, 1998, pp.107-117;
5. Freeman L.C. Centrality in social networks: conceptual clarification// Social Networks, 1, 1979, pp.215-239;
6. Freeman L.C. A set of measures of centrality based upon betweenness// Sociometry, 40, 1977, pp.35-41;
7. Freeman L.C., Borgatti S.P., White D.R. Centrality in valued graphs: A measure of betweenness based on network flow// Soc. Networks, 13, 1991, pp.141-154;
8. Katz L. A New Status Index Derived from Sociometric Index// Psychometrika, 1953, pp.39-43;
9. Newman M.E.J. A measure of betweenness centrality based on random walks// Soc. Networks, 27, 2005, pp.39-54;
10. Rochat Y. Closeness Centrality Extended to Unconnected Graphs: The Harmonic Centrality Index// ASNA, 2009;

INFLUENCE IN NETWORKS BY LONG-RANGE INTERACTIONS

Fuad Aleskerov, National Research University Higher School of Economics (HSE), V.A. Trapeznikov Institute of Control Sciences of Russian Academy of Sciences (ICS RAS), Moscow (alesk@hse.ru),
Natalia Meshcheryakova, National Research University Higher School of Economics (HSE), V.A. Trapeznikov Institute of Control Sciences of Russian Academy of Sciences (ICS RAS), Moscow (natamesc@gmail.com),
Sergey Shvydun, National Research University Higher School of Economics (HSE), V.A. Trapeznikov Institute of Control Sciences of Russian Academy of Sciences (ICS RAS), Moscow (shvydun@hse.ru).

Abstract: *A problem of key nodes detection in network structures is studied. We propose several models which take into consideration nodes attributes and the intensity of interactions. There are two different ideas on how to take into account long-range interactions between members of the network: a distance-based approach and an approach based on the idea of simulations. Our models help us to identify both explicit and hidden central elements which cannot be detected by classical centrality measures or other indices.*

Keywords: network analysis, influence indices, short and long-range interactions.