УДК 519.17

МЕТОД ОЦЕНКИ ВЛИЯНИЯ ВЕРШИН ПРИ АНАЛИЗЕ СЕТЕВЫХ СТРУКТУР

У.В. Александрова, В.А. Осипова

В статье предложен подход к анализу оценки влияния вершин при анализе сетей. Наряду с понятием центральности вершины рассматривается понятие центральности сети. Предлагается в качестве показателя оценки влияния вершины (актора) на топологию сети рассматривать понятие ключевой вершины, а также характеристику, близкую по своей структуре к индексу Банцафа.

The article proposes some approach to estimation of vertex influence in the analysis of networks. Along with the concept of centrality of the vertex, the concept of centrality of the network is considered. It is proposed to consider the concept of a key vertex as a measure of the impact of a vertex (actor) on the network topology, as well as a characteristic close in its structure to the Banzhaf index.

КЛЮЧЕВЫЕ СЛОВА

Анализ сетей, центральность вершины, центральность графа, ключевые вершины.

ДЛЯ ЦИТАТЫ

У.В. Александрова, В.А. Осипова. Метод оценки влияния вершин при анализе сетевых структур // Моделирование и анализ данных. 2019. №3. С. 32-36.

U.V. Alexandrova, V.A. Osipova Vertex influence estimation method in analyzing network structures. Modelirovaniye i analiz dannykh=Modelling and data analysis *(Russia)*. 2019, no.3, pp.32-36.

ВВЕДЕНИЕ

В последнее время задача анализа различных сложных сетей, в частности исследование их структуры и определение ключевых элементов, становится актуальной и занимает особое место наряду с другими исследованиями сетей. Топология сетей, отражающих взаимодействие вершин, оказывает влияние на принятие решений, связанных с их анализом. Сетевая топология представляется структурой в виде графа, вершинам (акторам, при исследовании социальных сетей) которого соответствуют узлы сети, а рёбрам — физические или информационные связи между вершинами. Различают различные виды сетевых топологий: полносвязная, неполносвязная (может применяться передача данных не напрямую между компьютерами, а через дополнительные узлы), элементарные: шина, звезда, кольцо.

Важной характеристикой сети является понятие центральности вершины. Различные определения центральности сводится к трем основным концепциям [1,5], рассмотренным ниже. Все они определяют некоторую меру, характеризующую важность актора. Наряду с понятием центральности вершины характеристикой сети выступает также понятие центральности самой сети.

В работе предлагается в качестве показателя оценки влияния вершины (актора) на топологию сети рассматривать понятие ключевой вершины, а также характеристику, близкую по своей структуре индексу к Банцафа, основанному на вычислении доли подграфов, в которых вершина является ключевой.

1. ДОСТАТОЧНАЯ ЦЕНТРАЛИЗОВАННОСТЬ СЕТИ И ОЦЕНКА ВЛИЯНИЯ ВЕШИНЫ (АКТОРА)

Рассмотрим граф $G = (V, E)$, где $V = \{v_1, .., v_n\}$ – множество вершин, а $E \subseteq V \times V$.

Рассмотрим три подхода к понятию центральности вершины [1, 5].

1. Центральность вершины как отражение потенциальной активности. Концепция этого подхода состоит в том, что наиболее активными считаются вершины, имеющие наибольшую степень. Центральность по степени $C_D(v_i)$ вершины v_i определяется как количество смежных с v_i вершин:

$$C_D(v_i) = \sum_{j=1}^{n} a_{ij}, \tag{1}$$

где $a_{ij} = 1$ тогда и только тогда, когда вершины соединены ребром. В работе [1] предложен вариант определения, не зависящий от размеров сети, - это относительная центральность по степени:

$$C'_D(v_i) = \frac{\sum_{j=1}^{n} a_{ij}}{n-1}.$$

2. Центральность по посредничеству вершины, как возможность управления передачей. Данный подход основывается на частоте прохождения через вершину кратчайших путей между всеми парами вершин. Предполагается, что соответствующий актор может влиять на группу акторов, поддерживая, придерживая или разрушая процесс передачи информации. Пусть g_{ij} – количество кратчайших путей от вершины графа v_i до вершины v_j, а $g_{ij}(k)$ – количество кратчайших путей от v_i до v_j, проходящих через v_k. Тогда мера центральности $C_B(v_k)$ для вершины v_k определяется следующим образом [1]:

$$C_B(v_k) = \sum_i \sum_j \frac{g_{ij}(k)}{g_{ij}} \qquad (i \neq j \neq k), \tag{2}$$

где ${}^{g_{ij(k)}}/_{g_{ij}} = 0$ тогда и только тогда, когда количество кратчайших путей между v_i и v_j равно нулю. Эта величина определяет долю кратчайших путей между вершинами v_i и v_j, проходящих через v_k, и ее можно интерпретировать как вероятность того, что случайно выбранный кратчайший путь между v_i и v_j пройдет через v_k. Меры (1) и (2) зависят от размеров сети, поэтому значения распределяются в большом диапазоне, что приводит к необходимости нормализации значений.

3. Центральность вершины как отражение ее независимости или эффективности. Эта концепция также базируется на идее контроля коммуникаций. Актор считается центральным до той степени, до которой он может избежать контроля со стороны других. На данный момент существует множество работ ([2],[3]), посвященных изучению данного подхода, но можно рассмотреть простой и естественный вариант такой меры, где центральность измеряется путем суммирования кратчайших расстояний от вершины до всех остальных:

$$C_C(v_k)^{-1} = \sum_i d(i, k). \tag{3}$$

Эту меру можно считать мерой удаленности, так как вершина будет иметь максимальную центральность в том случае, если она наиболее удалена от остальных вершин.

Наряду с понятием центральность вершины важной характеристикой сети является понятие центральность сети (графа). Считается, что индекс центральности графа, независимо от того, на какой мере центральности вершины он определен, должен обладать следующими свойствами [1]:

- показывать, до какой степени центральность самой центральной вершины превосходит центральность остальных вершин;
- выражаться отношением этого превосходства к максимально возможному значению для графа, содержащего рассматриваемое количество вершин.

Этот показатель характеризует разброс между такими индивидуальными показателями, как центральности вершин в графе.

В частности, если $C_G(v_i)$ −центральность вершины v_i графа $G(V,E)$, а $C_G(v^*)$ − максимальное значение центральности вершины, то индекс центральности графа можно определить как:

$$C_G = \frac{\sum_{i=1}^{n}[C_G(v^*) - C_G(v_i)]}{((n-1)(n-2))}, \quad (4)$$

Заметим, что $C_G = 0$ когда центральности всех вершин равны, $C_G = 1$ когда вершина v^* доминирует над всеми остальными вершинами.

Для дальнейшего анализа топологии сети предложим еще один способ оценки структуры графа и степени влияния вершины (актора) на структуру графа.

Определение 1. Граф G считается достаточно централизованным относительно заданного порогового значения q, если индекс центральности этого графа больше q: $C_G > q$, $0 \leq q \leq 1$.

Определение 2. Вершина v_i называется ключевой в достаточно централизованном относительно заданного порогового значения q графе G , если после ее удаления граф $G' = G\backslash\{v_i\}$ не является достаточно централизованным относительно заданного порогового значения q, т.е. $C_{G'} \leq q$.

Для иллюстрации этих понятий рассмотрим несколько примеров. Для значений центральности вершины воспользуемся концепцией центральности вершины как отражения потенциальной активности, т.е. центральности вершины по степени.
На Рис.1 изображена сетевая топология кольцо с $n > 3$ вершинами.

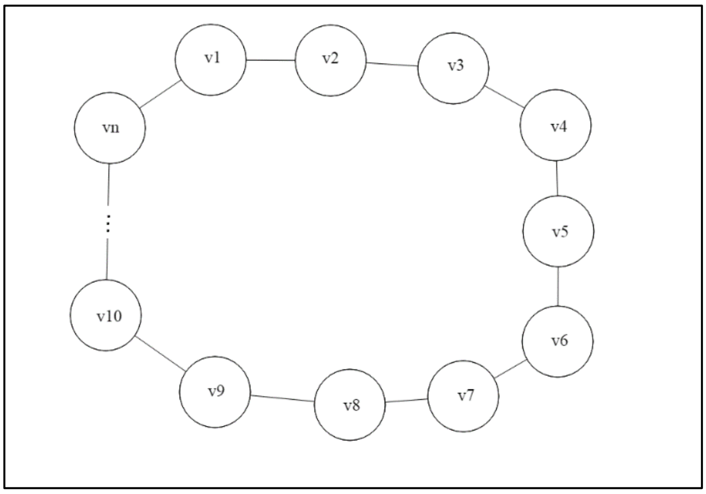

Рис.1 Граф кольцо

Тогда $C_D(v_1) = C_D(v_2) = \cdots = C_D(v_n) = 2$, индекс центральности графа $C_G = 0$. Граф не является достаточно централизованным относительно любого порогового значения q.

На Рис. 2 изображен графы топологии звезда с $n > 3$ вершинами:

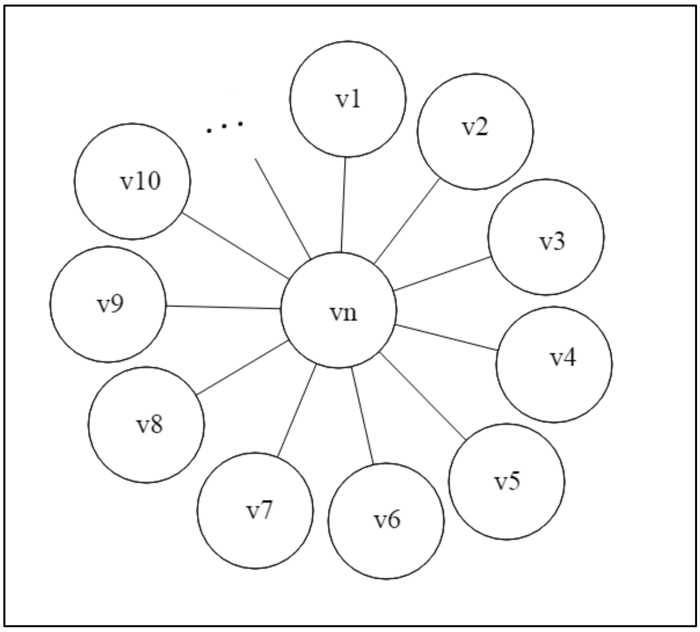

Рис.2 Граф звезда

Центральности вершин $v_1, \dots, v_{n-1}$ равны:

$$C_D(v_1) = C_D(v_2) = \cdots = C_D(v_{n-1}) = 1,$$
$$C_D(v_n) = n - 1.$$

Индекс центральности графа $C_G = 1$. Граф является достаточно централизованным относительно любого значения q. Вершина v_n является ключевой, поскольку для $G' = G \backslash \{v_n\}$ имеем $C_{G'} = 0$.

На рис.3 изображен граф смешанной топологии сети с $n = 12$ вершинами:

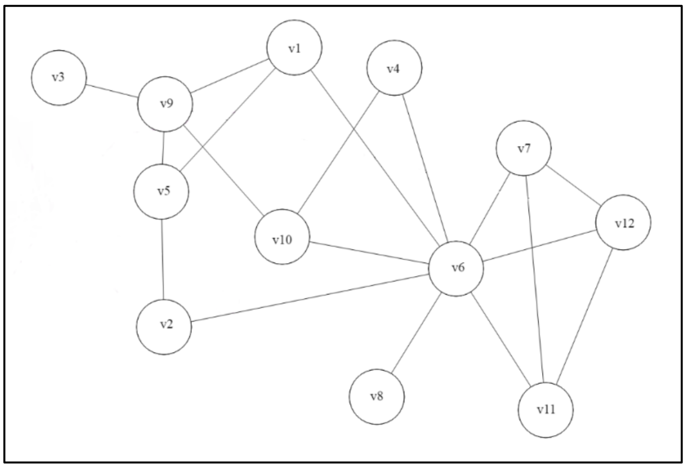

Рис.3 Граф смешанной топологии сети с $n = 12$ вершинами

Центральности вершин равны соответственно

$$C_C(v_1) = 3;\ C_C(v_2) = 2; C_C(v_3) = 1; C_C(v_4) = 2;\ C_C(v_5) = 3;$$
$$C_C(v_6) = 8; C_C(v_7) = 3;\ C_C(v_8) = 1;\ \ C_C(v_9) = 4; C_C(v_{10}) = 3;$$
$$C_C(v_{11}) = 3;\ C_C(v_{12}) = 3;$$

индекс центральности этого графа $C_G = 0{,}55$.

Отметим, что, например, относительно порогового значения $q = 0{,}8$ рассматриваемый граф не является достаточно централизованным, так как $C_G < 0{,}8$, а относительно порогового значения $q = 0{,}5$ граф является достаточно централизованным, т.к. $C_G > 0{,}5$ и ключевой вершиной является вершина v_6, поскольку для графа $G' = G\backslash\{v_6\}$ имеем $C_{G'} = 0{,}2777$.

В качестве показателя оценки влияния вершины (актора) на топологию сети рассматриваем характеристику, аналогичную по своей структуре индексу Банцафа [4], основанному на вычислении доли подграфов графа G, в которых вершина v_i является ключевой, т.е. отношению числа подграфов графа G, в которых вершина v_i является ключевой к общему числу подграфов графа G, содержащих вершину v_i:

$$\beta(v_i) = \frac{b_i}{N(b_i)},$$

где b_i – число подграфов графа G, в которых вершина v_i является ключевой, а $N(b_i)$ – общее число подграфов графа G, содержащих вершину v_i.

Алгоритмы определения индекса центральности графов и оценки влияния вершины на топологию сети имеют вычислительную сложность порядка ($O(n^3)$), что усложняет их использование в больших реальных сетях.

ЛИТЕРАТУРА

1. Freeman L. C. Centrality in social networks. Conceptual clarification // Soc. Networks. 1978/79. V. 1. P. 215-239.
2. Leavitt H. J. Some effects of certain communication patterns on group performance // J. Abnorm. and Soc. Psychology. 1951. V. 46. P. 38-50.
3. Bavelas A. A mathematical model for group structure // Human Organization. 1948. V. 7. P. 16-30.
4. Алескеров Ф.Т., Хабина Э.Л., Шварц Д.А. Бинарные отношения, графы и коллективные решения. – М.: ФИЗМАТЛИТ, 2017. – 344 с.
5. Щербакова Н.Г. Меры центральности в сетях. – Институт вычислительной математики и математической геофизики СО РАН,630090, Новосибирск, 2015. – 13 с.

Работа поступила 07.04.2019г.