АКСИОМАТИКА ЦЕНТРАЛЬНОСТИ В КОМПЛЕКСНЫХ СЕТЯХ

Н. Г. Щербакова
Институт вычислительной математики и математической геофизики СО РАН,
630090, Новосибирск, Россия

УДК 001.12+303.2

Рассмотрены системы аксиом, определяющих базовые свойства меры центральности. Результаты анализа классических мер на соответствие аксиомам позволяют судить о способе выявления важных узлов сети.

Ключевые слова: меры центральности сетевых узлов, аксиомы центральности, анализ мер центральности.

Systems of the axioms that define basic properties of a centrality measure are considered. Analyzing whether a given centrality measure satisfies the axioms is a way to estimate the method to account for the importance of the nodes of a network.

Key words: node centrality measures, axioms for centrality, centrality measures analysis.

Введение. Центральность — это категория, определяющая роль актора в сети. Решается задача идентификации ключевых акторов, которые структурно хорошо интегрированы в сеть и влияют на процессы, протекающие в сети. Сеть в этом случае представляется в виде графа, вершины которого соответствуют акторам, а ребра — связям между ними. Многообразие предложенных мер центральности, прежде всего основанных на интуитивном представлении, свидетельствует о неоднозначности подхода к вопросу. Требуется определить, что общего имеют меры центральности, какие структурные свойства узлов сети заложены в определение центральности и существуют ли структурные свойства, которые не измеряются с помощью центральности. Сравнение мер центральности является трудной задачей по многим причинам. Одна из них — недостаточный объем исходных данных, на основе которого трудно сделать правильные выводы. Некоторые авторы считают, что разные меры выявляют разные аспекты центральности, и нет смысла их сравнивать [1]. Тем не менее, поведение мер центральности и определение, действительно ли рассматриваемая мера соответствует своему предназначению, представляют предмет для изучения. В данной работе рассматриваются аксиоматические подходы к формулировке базовых требований, предъявляемых к мере центральности.

1. Аксиоматика Sabidussi. В работе [2] предложен ряд критериев, которым должна удовлетворять функция, которая считается мерой центральности вершин графа. В частности, мера должна адекватно описывать центр графа, так как автор считает, что индекс центральности графа следует определять на основе меры центральности вершин.

Обозначим через S класс всех конечных связных неориентированных графов $G(V,\ E)$, а через S_n — класс таких графов с n вершинами ($|V| = n$). Обозначим через U любой класс неориентированных графов, а через U_n — класс таких графов с n вершинами ($n \in N$). Предположим, что на S_n задана функция σ, которая сопоставляет с графом $G \in S_n$ вектор $(a_1, a_2, \ldots, a_n)$ неотрицательных вещественных чисел. Если $v_1, v_2, \ldots, v_n$ — вершины графа G, то значение a_i соответствует вершине v_i и обозначается как $\sigma_G(v_i)(i = 1{,}2{,}...n.)$

Центр графа $C_\sigma(G)$ относительно функции σ определяется как

$$C_\sigma(G) = \{v \in V : \sigma_G(v) \leqslant \sigma_G(x), \forall x \in V\}.$$

Без наложения ограничений на функцию σ слово „центр" — это просто название для множества вершин. Заметим, что если σ — мера центральности, то вершина, принадлежащая центру графа, имеет наименьшее значение центральности. Автор в работе [2] определил меру „центральность по близости" на основе дистанций между вершинами, а затем обобщил подход в виде аксиоматики.

Определим понятия добавления ребра и допустимого переключения ребра, которые понадобятся при формулировке аксиом. Пусть в графе G имеются две различные вершины v_1, v_2, такие, что $e = (v_1,v_2) \notin E$, образуем множество $E_1 = E \cup \{\mathrm{e}\}$, тогда говорим, что граф $G_1(V,\ E_1)$ получен из G добавлением ребра, инцидентного v_1 (v_2). Теперь пусть v, v_0, v_1 — различные вершины графа $G(V,\ E)$. Пусть $e_0 = (v,v_0) \in E$, а $e_1 = (v,v_1) \notin E$. Образуем новый граф удалением ребра e_0 и добавлением ребра e_1 и назовем эту процедуру *переключением* с ребра e_0 на e_1. Переключение назовем *допустимым*, если выполняется условие $d(v_0,\ v_1) \leqslant d(v_0,\ v)$, где $d(v_1,\ v_2)$ — расстояние между вершинами v_1, v_2 (длина кратчайшего пути). При допустимом переключении „центр поворота" v не лежит между вершинами v_0, v_1, и граф остается связным. Далее под „термином" переключение будем понимать допустимое переключение.

Пусть имеются класс графов U и функция σ на S_n. Класс U_n называется *σ-допустимым*, а σ — *мерой центральности вершин для класса графов U_n* тогда и только тогда, когда выполняются условия:

(S1) Класс U_n замкнут относительно изоморфизма, т. е. если $G \in U_n$ и η — изоморфизм $G \to H$, то $H \in U_n$. Изоморфизмом графов называют отображение η множества вершин графа G на множество вершин графа $H : V_G \to V_H$, такое, что

$$\forall(v_1,v_2) \in E_G \to (\eta(v_1),\eta(v_2)) \in E_H,$$

$$\forall(v_1,v_2) \in E_H \to \left(\eta^{-1}(v_1),\eta^{-1}(v_2)\right) \in E_G,$$

где η^{-1} — обратное отображение. В случае ориентированного графа (орграфа) сохраняется ориентация, в случае взвешенного — вес. Таким образом, изоморфные графы имеют одинаковую структуру.

(S2) Если $G \in U_n$ & $v \in C_\sigma(G)$, а граф H получен переключением ребра на v или добавлением ребра, инцидентного v, то $H \in U_n$. Будем говорить, что класс U_n замкнут относительно переключения или добавления ребра к центральной вершине.

(S3) Если $G \in U_n$ и η — изоморфизм $G \to H$, то

$$(\forall v \in V_G)\, \sigma_H\left(\eta(v)\right) = \sigma_G(v).$$

(S4) Если $G \in U_n$ & $v \in V_G$, а $H \in U_n$ получен из G добавлением ребра, инцидентного v, то

$$\sigma_H(v) < \sigma_G(v) \& \left((\forall h \in V_H)\sigma_H(h) \leqslant \sigma_G(h)\right).$$

(S5) Если $G \in U_n$ & $v \in C_\sigma(G)$, а H получен из G переключением ребра на v или добавлением ребра, инцидентного v, то

$$\sigma_H(v) < \sigma_G(v) \ \& \ v \in C_\sigma(H),$$

т. е. центр укрепляет свою позицию, когда к нему подключился новый канал связи.

Если σ — функция центральности вершины для класса U_n и $G \in U_n$ & $v \in V_G$, то неотрицательное число $\sigma_G\ (v)$ называется центральностью вершины v в графе G.

На основании формального определения меры центральности вершины вводятся также формальные определения неравенства графов $G \underset{\sigma}{<} H$ (граф G менее σ-централизован, чем H) и „индекса центральности“ для класса графов.

Таким образом, охарактеризован класс графов, для которого вводится мера центральности, а основными являются аксиомы монотонности S4–S5, описывающие поведение меры при присоединении или переключении ребра к любой или к центральной вершине.

Однако показано, что некоторые естественные классы незамкнуты относительно переключения/добавления ребра к центральной вершине (S2), например класс деревьев. Кроме того, как замечено в работе [2], трудно, например, доказать, что центральность по близости (C_C), классическое определение которой дано именно в этой работе, является мерой центральности для класса всех конечных связных графов (сложность проверки S5). Также показано, что такая естественная мера центральности вершины графа G, как $m(x) = \max\limits_{v \in G} d(x,v)$, не является σ-центральностью.

Таким образом, согласно предложенным аксиомам, известные меры центральности таковыми не являются. Более того, хотя аксиомы обеспечивают ряд желательных характеристик мер центральности, они не объясняют суть понятия центральности. Тем не менее, работа дала толчок к появлению других подходов к формализации понятия центральности.

2. Аксиоматика Nieminen. Подобная аксиоматика для конечных слабо связных орграфов (DG) приведена в работе [3]. Концепция имеет смысл только для „геометрической“ центральности, так как изложена в терминах расстояний между вершинами. А именно, для каждой вершины v графа $G(V,\ E)$ определяется вектор субординации (*subordinate vector*) $\mathbf{sv}(v) = (sv_1,\ sv_2,\ \dots sv_{n-1})$, sv_i— множество вершин, достижимых из v и находящихся на расстоянии i от v. В этих терминах условия центральности вершины $C(v)$ таковы:

(N1) $C(v)$ — целое число.

(N2) $C(v) = 0$ тогда и только тогда, когда у вершины v нет достижимых вершин.

(N3) Если $\mathbf{sv}(v_1)$ получен из $\mathbf{sv}(v)$ добавлением одной новой достижимой вершины, находящейся на любом расстоянии от v_1, то $C(v_1) > C(v)$.

(N4) Если $\mathbf{sv}(\mathrm{v}_1)$получен из $\mathbf{sv}(v)$ уменьшением расстояния до любой достижимой вершины, то $C(v_1) > C(v)$.

Таким образом, аксиомы монотонности N3–N4 формулируются в терминах увеличения количества достижимых вершин или уменьшения расстояния до них. Следует отметить, что, например, для центральности по близости (C_C), определенной в терминах дистанций в работе [2], аксиома (N4) не выполняется. Здесь неравенства направлены в другую сторону по сравнению с предыдущей работой, т. е. значение центральности увеличивается с уменьшением расстояния.

3. Свойства мер центральности вершин в неориентированных невзвешенных связных графах. В статье [4] продолжена работа по определению желаемых свойств для меры центральности вершин графа. Рассматривается класс неориентированных невзвешенных связных графов. Равенство значений меры центральности для изоморфных вершин (S3, Sabidussi) не выделено в отдельное свойство, но считается естественным, чтобы в случае идентичной сетевой структуры соответствующие вершины имели одно и то же значение меры центральности. Сформулированы следующие три свойства, касающиеся добавления ребра графа.

Свойство 1. *Монотонность по отношению к дистанции.* Если расстояние от вершины x до по крайней мере одной вершины y становится меньше за счет добавления одного ребра, то значение меры центральности для x увеличивается (это же верно для y ввиду неориентированности графа). В терминах обмена информацией актор имеет больший потенциал, если у него появится больше прямых контактов. Формально это выглядит так:

$$(V_G = V_{G'}; v, w, x, y \in V_G, v \neq w, x \neq y; (v, w) \notin E_G; E_{G'} = E_G \cup (v, w);\ \exists\ y\ d_{G'}(x,y) < d_G(x,y))$$
$$\to (\sigma_{G'}(x) > \sigma_G(x))\ \ \&\ \ (\sigma_{G'}(y) > \sigma_G(y))$$

Свойство 2. *Монотонность по отношению к количеству кратчайших путей.* Если количество кратчайших путей от вершины x до хотя бы одной вершины y увеличивается за счет добавления ребра, то значение меры центральности для x увеличивается (то же для y). Чем больше прямых контактов актора с другими членами группы, тем более достоверна передаваемая информация. Формально это выглядит так:

$$\Big(V_G = V_{G'}; v, w, x, y \in V_G, v \neq w, x \neq y; (v, w) \notin E_G; E_{G'} = E_G \cup (v, w); d_{G'}^{(v,w)}(x,y) = d_G(x,y)\Big)$$
$$\to (\sigma_{G'}(x) > \sigma_G(x))\ \ \&\ \ (\sigma_{G'}(y) > \sigma_G(y))\,.$$

Здесь $d_{G'}^{(v,w)}(x,y)$ — длина кратчайшего пути от x до y, проходящего через ребро $(v,\ w)$. Данный путь, ввиду предпосылки, появляется только после добавления ребра $(v,\ w)$, соответственно, количество кратчайших путей увеличивается.

Свойство 3. *Влияние на ранжирование.* С добавлением ребра между вершинами x и y ранжирование этих вершин не изменится. Предполагается, что если актор x был лучше интегрирован в сеть, чем y, то добавление прямого контакта не изменит это положение. Формально это выглядит так:

$$(V_G = V_{G'};\quad x, y \in V_G, x \neq y;\quad E_{G'} = E_G \cup (x, y)) \to (\sigma_{G'}(x) \geqslant \sigma_{G'}(y))\,;$$
$$(\sigma_G(x) = \sigma_G(y)) \to (\sigma_{G'}(x) = \sigma_{G'}(y))\,.$$

Были проанализированы известные меры центральности на соответствие свойствам 1–3. В табл. 1 приведены результаты анализа, здесь „+“ означает, что мера центральности обладает данным свойством, „—“ — не обладает. Поскольку в формулировках даже базовых мер центральности, приводимых различными авторами, имеются нюансы, приведем формулы для вычисления мер.

Центральность по степени (*degree centrality*) [5]:

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

где a_{ij}— элемент матрицы смежности графа.

Таблица 1

Свойства и меры центральности

Мера центральности	Свойство 1	Свойство 2	Свойство 3
Центральность по степени [5]	—	—	—
Центральность по посредничеству [6]	—	—	—
Центральность по близости [6]	+	—	—
Центральность собственного вектора [7]	—	—	—
Центральность по Кацу [8]	+	+	+

Центральность по посредничеству (*betweenness centrality*) [6]:

$$C_B(k) = \sum_i \sum_j \frac{g_{ij}(k)}{g_{ij}} \left(i \neq j \neq k\right),$$

где g_{ij} — количество кратчайших путей от вершины i до вершины j графа, а $g_{ij}(k)$ — количество кратчайших путей от i до j, проходящих через k.

Центральность по близости (*closeness centrality*) [6]:

$$C_C(k) = \frac{1}{\sum\limits_j d(k,j)},$$

где $d(k,\, j)$ — кратчайшее (геодезическое) расстояние от вершины k до вершины j.

Центральность собственного вектора (*eigenvector centrality*) [7]:

$$C_E(i) = e_i = \frac{1}{\lambda_{\max}} \sum_{j=1}^{n} a_{ji} e_j,$$

где $\mathbf{e} = (e_1,\, \ldots,\, \mathrm{e}_n)^\top$ — главный собственный вектор матрицы смежности A, соответствующий максимальному собственному значению $\lambda_{\max}$.

Таким образом, центральность вершины i пропорциональна сумме центральностей соседей, причем статус вершины i зависит от статуса вершин j, таких, что ребро $j \to i$ принадлежит графу (в работе [7] определение вводится для случая ориентированного графа).

Центральность по Кацу (*Katz's centrality*) [8]:

$$C_K(x) = \mathbf{1}^\top (\sum_{i=1}^{\infty} b^i A^i) \mathbf{e}_x,$$

где $\mathbf{1} = (1{,}1,\, \ldots{,}1)^\top$ — вектор, состоящий из единиц, $\mathbf{e}_x = (0,\, 0,\, \ldots,\, 1,\, 0{,}\ldots,\, 0)^\top$ — единичный вектор, b — произвольный (обычно положительный) взвешивающий фактор.

Для центральности по степени свойства 1 и 2 не имеют смысла, поскольку при ее вычислении учитываются только прямые контакты. Выполнение свойства 3 очевидно. Центральность по близости рассматривается в значении, обратном к сумме расстояний до всех вершин графа, так, чтобы уменьшение расстояния увеличивало значение центральности.

Заметим, что если мера не обладает свойством, достаточно привести контрпример, что и сделано в работе [4]; выполнение свойства требует доказательства. Для центральности по близости выполнение свойства 1 является очевидным. Выполнение свойств 1, 2 для центральности по Кацу представляется очевидным, так как увеличивается взаимосвязанность вершин; выполнение свойства 3 подтверждено путем моделирования.

Таким образом, известные меры центральности не обладают даже такими интуитивно обоснованными свойствами, что сказывается и на анализе роли акторов. Отсюда следует, что свойства можно рассматривать как характеристики мер, и необходимо анализировать требования к приложению, прежде чем использовать определенную меру центральности. Предложенные свойства рассмотрены при наличии определенных ограничений: графы невзвешенные неориентированные; не учитываются реальная интенсивность взаимодействия актора с новыми членами и ослабление уже имеющихся связей. Поэтому требуется рассмотрение свойств и поведения мер центральности в более широком контексте, а также разработка соответствующих сценариев, позволяющих применять меры для анализа акторов в рамках конкретных приложений.

4. Аксиоматика Boldi — Vigna. Авторы работы [9] так охарактеризовали меру центральности: она должна быть, с одной стороны, устойчивой (применима к произвольным орграфам, возможно, несвязным), с другой — понятно сформулирована (имела прозрачную комбинаторную интерпретацию). К аксиомам авторы предъявляют несколько требований: они должны иметь ясную семантику, их выполнимость может быть проверена для большинства мер центральности, они должны быть сформулированы, избегая ловушек небольшого ограниченного множества данных, для которого многие меры центральности могут показать отрицательные результаты (т. е. должны использоваться асимптотические определения). Предполагается, что рассматриваемые меры инвариантны относительно изоморфизма, т. е. зависят только от структуры графа, а не от конкретной нумерации вершин.

Изучается реакция мер центральности на изменение размера, изменение (локальной) плотности и добавление дуги. Ожидается, что узлы сети, принадлежащие к группам большего размера, при других фиксированных параметрах должны быть более важными, а узлы с более плотным соседством (например, соответствующих акторам, имеющим больше друзей) при других фиксированных параметрах должны быть более важными. Сеть представляется в виде ориентированного графа. Заметим, что первоначально меры центральности формулировались в терминах неориентированных связных графов, в дальнейшем они были переформулированы так, чтобы их можно было использовать и для несвязных орграфов. Здесь рассматриваются пути в графе, оканчивающиеся в вершине, для которой вычисляется значение центральности.

Для проверки выполнимости свойств требуется вычисление всех рассматриваемых мер для всех узлов сети. Удачным приемом сокращения вычислений является рассмотрение класса сильно связных вершинно-транзитивных графов (для любых двух вершин x и y существует автоморфизм, отображающий x на y). Такие графы можно считать базовыми строительными блоками: они проявляют в высокой степени симметрию, что облегчает вычисления. Так как сравнивается плотность, естественно рассмотреть наиболее плотный сильно связный вершинно-транзитивный граф — клику (любые две вершины связаны дугой) и наименее плотный такой граф — направленный цикл. Тем более, что k-клики и p-циклы существуют для любых k и p.

Изучению подлежат следующие вопросы. Все вершины k-клики имеют одинаковое значение для любой меры центральности, то же верно для вершин p-цикла, но какие вершины важнее? Практически очевидно, что для $k = p$ элементы клики важнее, действительно, это подтверждается вычислением почти всех мер центральности. Каково влияние размера графа на значение меры центральности?

4.1. *Аксиомы центральности.*

В1 (*аксиома размера*). Рассмотрим граф $S_{k,p}$, состоящий из двух компонент: k-клики и p-цикла. Мера центральности удовлетворяет аксиоме размера, если для каждого k существует число P_k, такое, что для всех $p \geqslant P_k$ в графе $S_{k,p}$ центральность вершины в p-цикле строго больше центральности вершины в k-клике, а для каждого p существует K_p, такое, что для всех $k \geqslant K_p$ центральность вершины в k-клике строго больше центральности вершины в p-цикле.

Теперь рассмотрим граф, состоящий из направленного k-цикла и направленного p-цикла, соединенных двунаправленным мостом $x \leftrightarrow y$; вершина x принадлежит k-циклу, вершина y принадлежит p-циклу. Если $k = p$, вершины x и y симметричны и должны иметь одинаковое значение центральности. Далее, наращивая плотность k-цикла, постепенно превращаем его в k-клику. Это изменение плотности локально относительно x, степень y не изменяется.

В2 (*аксиома плотности*). Рассмотрим граф $D_{k,p}$, состоящий из k-клики и p-цикла ($k,\, p \geqslant 3$), связанных двунаправленным мостом $x \leftrightarrow y$, где x является вершиной клики, а y — вершиной цикла. Мера центральности удовлетворяет аксиоме плотности, если для $k = p$ центральность x строго больше центральности y.

Заметим, что эта аксиома формулируется для $k = p$, т. е. размер не вовлечен в формулировку. При доказательстве факта выполнимости аксиомы плотности для конкретной меры k и p выступают как независимые параметры; выявляется, при каком значении k (определенном как функция от p) значение центральности вершины x становится больше значения центральности для вершины y (если становится). Этот водораздел может дать представление о том, насколько плохо для меры центральности выполняется аксиома.

В3 (*аксиома монотонности*). Мера центральности удовлетворяет аксиоме монотонности, если для каждого графа G и каждой пары вершин x и y, таких, что ребро $x \to y$ не принадлежит E_G, если добавить в G такое ребро, то центральность y возрастет.

Эта аксиома по существу выполняется для всех мер центральности в случае рассмотрения сильно связных графов. Заметим, что аксиома сформулирована в понятиях изменения значения меры центральности, а не в понятиях неуменьшения ранга y. Дело в том, что такая слабая аксиома монотонности выполняется даже для меры центральности, сопоставляющей с каждым узлом любой сети константу, что делает ее неинтересной для рассмотрения. Однако, возможно, существуют меры центральности, при применении которых ранг y даже уменьшится, что представляет интерес.

4.2. *Определения мер центральности.* Остановимся кратко на вариантах определения мер центральности, для которых в работе [9] приведено доказательство о выполнимости или невыполнимости аксиом В1–В3.

4.2.1. Геометрические меры.

Центральность по входящей степени (*in-degree centrality*): количество ребер, входящих в вершину x, $\;C_D(x) = \deg^-(x)$.

Центральность по близости (*closeness centrality*) [10] в применении к орграфам, не являющимся сильно связными (если из вершины y вершина x недостижима, то $d(x,\, y) = \infty$, поэтому такие вершины исключаются из рассмотрения):

$$C_C(x) = \frac{1}{\sum\limits_{d(x,y)<\infty} d(x,y)}.$$

Индекс Лин (*Lin's index*) [11] — корректировка центральности по близости для случая несвязных графов с непустым множеством взаимно достижимых вершин путем взвешивания квадратом количества таких вершин:

$$C_L(x) = \frac{|\{y|d(y,x) < \infty\}|^2}{\sum\limits_{d(x,y)<\infty} d(x,y)}.$$

Гармоническая центральность (*harmonic centrality*) [12], сформулированная на основе определения гармонического среднего (*harmonic mean*) [13], это корректировка меры с учетом недостижимых вершин, вносящих значение 0 в сумму, в предположении, что $\infty^{(-1)} = 0$:

$$C_{HC}(x) = \sum_{y \neq x} \frac{1}{d(y,x)} = \sum_{d(y,x)<\infty, y\neq x} \frac{1}{d(y,x)}.$$

4.2.2. Спектральные меры.

Левый главный собственный вектор матрицы смежности — вектор $\mathbf{e}$, являющийся решением уравнения $\mathbf{e}\mathrm{A} = \lambda\mathbf{e}$, причем это вектор, соответствующий максимальному значению λ.

Индекс Seeley (*Seeley's index*) [14] — главный собственный вектор нормализованной по строкам матрицы смежности (неотрицательная матрица смежности с ненулевыми строками нормализуется так, что все строки имеют одинаковую l_1-норму; достигается, например, делением элементов строки на сумму по строке). В матричном виде — это вектор $\mathbf{v}$, являющийся решением уравнения $\mathbf{v}\bar{A} = \mathbf{v}$.

Центральность по Кацу (*Kat'z centrality*) [8]:

$$C_K = \mathbf{1}\sum_{i=0}^{\infty} b^i A^i = \mathbf{1}\left(I - bA\right)^{-1}.$$

Для сходимости суммы фактор затухания b должен быть меньше, чем $1/\lambda$, где λ — наибольшее собственное значение матрицы A.

PageRank [15] — это уникальный вектор $\mathbf{p}$:

$$\mathbf{p} = \alpha\mathbf{p}\bar{A} + (1-\alpha)\mathbf{v},$$

где $\bar{A}$ — l_1-нормированная матрица смежности, $\alpha \in [0,1)$ — коэффициент затухания, $\mathbf{v}$ — вектор предпочтения. Гарантию сходимости дает замена нулевых строк на вектор $\mathbf{v}$. В работе [9] предлагается вариант определения без дополнительного преобразования матрицы $\bar{A}$:

$$\mathbf{p} = (1-\alpha)\mathbf{v}(I-\alpha\bar{A})^{-1}.$$

HITS (*hyperlink-induced topic search*) [16] — процесс одновременного вычисления двух типов центральности: *authority centrality*, определяющей „центры влияния“ (*authoritative pages*, a), — страницы, на которые указывают влиятельные центры внимания, и *hub centrality*, определяющей „центры внимания“ (*hubs*, h), которые указывают на влиятельные страницы. Итеративный процесс стартует с $a_0 = 1$, правило вычисления a_{i+1} и h_{i+1}:

$$h_{i+1} = a_i A^\top,$$

$$a_{i+1} = h_{i+1} A.$$

Процесс сходится к вычислению левого главного собственного вектора матрицы $A^\top A$, задающего значения центральности для вершин как центров влияния (*authority centrality*), и левого главного собственного вектора матрицы $AA^\top$, задающего значения центральности для вершин как центров внимания (*hub centrality*). В работе [16] сделано важное замечание, касающееся произведения матриц. Элемент (i, j) матрицы $A^\top A$ задает количество страниц, ссылающихся одновременно на страницы p_i и p_j; элемент (i, j) матрицы AA^T задает количество страниц, на которые ссылаются одновременно страницы p_i и $p_{j.}$ Если обратиться к понятиям коцитирования [17], [18] и библиографического сочетания публикаций [19], используемым в библиометрии, то в работе [20] показано, что элемент (i, j) матрицы $A^\top A$ задает количество публикаций, которые ссылаются (цитируют) одновременно на публикации d_i и d_j (коэффициент коцитирования), а элемент (i, j) матрицы $AA^\top$ задает количество публикаций, на которые ссылаются одновременно публикации d_i и d_j (коэффициент библиографического сочетания).

SALSA [21]: подход к определению двух типов центральности аналогичен *HITS*, но матрицы A и $A^\top$ l_1-нормализованы. Итеративный процесс стартует с $a_0 = 1$, правило вычисления a_{i+1} и h_{i+1}:

$$h_{i+1} = a_i \overline{A^\top},$$

$$a_{i+1} = h_{i+1} \overline{A}.$$

4.2.3. Меры, основанные на подсчете количества путей.

Центральность по посредничеству (*betweenness centrality*) [22], сформулированная для случая, когда орграф не является сильно связным:

$$C_B(x) = \sum_{y,z \neq x, g_{yz} \neq 0} \frac{g_{yz}(x)}{g_{yz}},$$

где g_{yz} — количество кратчайших путей от вершины y до вершины z графа, $g_{yz}(x)$ — количество кратчайших путей от y до z, проходящих через x.

Авторы работы [9] отмечают, что все спектральные меры могут быть интерпретированы как основанные на подсчете путей, так как они зависят от предела некоторых сумм степеней матрицы смежности A или от предела степеней A, и эти алгебраические операции могут быть выражены в терминах путей. Например, при нахождении главного собственного вектора неотрицательной матрицы может использоваться степенной метод, вычисляющий предел выражения $\mathbf{1}A^k/||\mathbf{1}A^k||$ при $k \to \infty$.

Таблица 2

Аксиомы Boldi — Vigna и меры центральности

Мера центральности	A1. Размер	A2. Плотность	A3. Монотонность
Входящая степень вершины (*in-degree*)	только k	+	+
Гармоническая центральность [12]	+	+	+
Центральность по близости [10]	—	—	—
Lin's index [11]	только k	—	—
Центральность по посредничеству [22]	только p	—	—
Центральность собственного вектора	только k	+	—
Seeley's index [14]	—	+	—
Центральность по Кацу [8]	только k	+	+
PageRank [15]	—	+	+
HITS [16]	только k	+	—
SALSA [21]	—	+	—

В свою очередь, $\mathbf{1}A^k$ — это вектор, ассоциирующий с каждой вершиной количество путей длиной k, оканчивающихся в вершине, а главный собственный вектор выражает относительный рост количества путей, оканчивающихся в вершине по мере роста длины путей.

4.3. *Результаты проверки выполнимости аксиом.* В табл. 2 приведены результаты проверки выполнимости аксиом B1–B3 для мер центральности. Выражение „только к“ означает, что выполняется только часть аксиомы B1, а именно: „для каждого p существует K_p, такое, что для всех $k \geqslant K_p$ центральность вершины в k-клике строго больше центральности вершины в p-цикле“, соответственно, „только p“ означает выполнение только части „для каждого k существует число P_k, такое, что для всех $p \geqslant P_k$ в графе $S_{k,p}$ центральность вершины в p-цикле строго больше центральности вершины в k-клике“.

Из табл. 2 видно, что только для гармонической центральности верны все аксиомы. Все спектральные меры чувствительны к локальной плотности графа. Нормализованные меры (*Seeley's index*, *PageRank*, *SALSA*) нечувствительны к изменению размера, остальные, кроме центральности по посредничеству, чувствительны только к увеличению параметра k. Центральность по близости показала худшие результаты, а гармоническая центральность — лучшие. Авторы работы [9] считают, что геометрические меры, несмотря на простоту, показывают хорошие результаты не только для социальных сетей, но и для систем информационного поиска.

Заключение. В настоящей работе рассмотрены различные подходы к формализации понятия центральности узла на основе аксиом. Приведенные аксиомы затрагивают только некоторые аспекты поведения мер центральности, но они позволяют лучше понять существующие подходы к определению важных узлов. Все авторы проанализированных статей предполагают, что при изоморфизме сетевой структуры соответствующие вершины должны иметь одно и то же значение меры центральности. Основными аксиомами, сформулированными в этих статьях, являются аксиомы монотонности, касающиеся поведения меры в случае изменения одного из параметров графа.

В пионерской работе [2] сформулировано понятие меры центральности вершины и центра графа, принадлежащего классу конечных неориентированных графов с n вершинами. Наиболее важные аксиомы характеризуют изменение значения меры центральности вершины при добавлении ребра, инцидентного вершине, в случае, если это произвольная вершина или вершина, принадлежащая центру графа. Направление неравенств в формулировке аксиом говорит о том, что все-таки имелась в виду центральность по близости, определенная в этой работе, хотя для нее не доказана выполнимость аксиомы S5.

Желаемые свойства меры центральности вершин слабо связных орграфов приведены в работе [3]. Свойства сформулированы в терминах расстояния до достижимых вершин, поэтому подходят только для геометрических мер центральности. Аксиомы монотонности касаются добавления новой достижимой вершины или уменьшения расстояния до достижимой вершины.

Авторы более поздней работы [4] рассматривали класс конечных неориентированных невзвешенных связных графов. Свойства монотонности сформулированы не только в терминах уменьшения расстояния от рассматриваемой вершины до других вершин, но и в терминах увеличения количества кратчайших путей от вершины и изменения ранга вершин. Однако, как показано в работе, из рассмотренных пяти базовых мер центральности всеми свойствами обладает только мера центральности по Кацу.

И, наконец, в работе [9] представлены три равнозначимых аксиомы, характеризующих изменения различных параметров произвольных орграфов. А именно, рассмотрено, как значение меры может зависеть от размеров графа, локальной плотности и добавления ребра. Проанализированы одиннадцать мер центральности, в том числе редко встречающиеся в работах других авторов, такие как *Seeley's index*, *HITS*, *SALSA*. Однако и в данном случае все аксиомы выполняются только для гармонической центральности.

Таким образом, одним набором аксиом не удается охватить не только все известные меры центральности, но даже и определенный тип центральности, например, меры центральности геометрического типа. И это несмотря на то, что формулируются достаточно общие и разумные требования к мере центральности. В этом смысле представляют интерес работы, которые определяют аксиомы, характеризующие конкретную меру центральности. Так, в работе [23] рассматривается класс сильно связных орграфов. В терминах функции ранжирования сформулированы пять аксиом, определяющих необходимые и достаточные условия, при выполнении которых функция ранжирования является *PageRank*. В работе [24] рассмотрены необходимые и достаточные условия для меры ранжирования, основанной на главном собственном векторе неприводимой неотрицательной матрицы.

Список литературы

1. Friedkin N. E. Theoretical foundations for centrality measures // Am. J. of Sociology. 1991. V. 96, iss. 6. P. 1478–1504.

2. Sabidussi G. The centrality index of a graph // Psychometrika. 1966. V. 31, iss. 4. P. 581–603.

3. Nieminen U. J. On the centrality in a directed graph // Social Sci. Research. 1973. V. 2, iss. 4. P. 371–378.

4. Langherr A., Friedl B., Heidemann J. A critical review of centrality measures in social networks // Business & Information Systems Engineering. 2010. V. 2, iss. 6. P. 371–385.

5. Nieminen U. J. On centrality in a graph // Scandinavian J. of Psychology. 1974. V. 15. P. 322–336.

6. Freeman L. C. Centrality in social networks. Conceptual clarification // Social networks. 1978/79. V. 1. P. 215–239.

7. Bonacich P., Lloyd P. Eigenvector-like measures of centrality for asymmetric relations // Social Networks. 2001. V. 23, iss. 3. P. 91–201.

8. Katz L. A new index derived from sociometric data analysis // Psychometrika. 1953. V. 18. P. 39–43.

9. Boldi P., Vigna S. Axioms for centrality. Cornell University Library. [Electron. resource]. http://arxiv.org/abs/1308.2140.

10. Bavelas A. Communication patterns in task-oriented groups // J. of the Acoustical Society of America. 1950. V. 22. P. 271–288.

11. Lin N. Foundations of Social Research. NY: McGraw-Hill, 1976.

12. Rochat Y. Closeness centrality extended to unconnected graphs: The harmonic centrality index // Proc. of the Applications of Social Network Analysis (ASNA 2009), Zurich, Aug. 26–28. [Electron. resource]. http://infoscience.epfl.ch/record/200525/files/[EN]ASNA09.pdf.

13. Marchiori M., Latora V. Harmony in the small-world // Physica A: Statistical Mechanics and its Applications. 2000. V. 285, iss. 3–4. P. 539–546.

14. Seeley J. R. The net of reciprocal influence: A problem in treating sociometric data // Canadian J. of Psychology. 1949. V. 3. P. 234–240.

15. Brin S., Page L. The anatomy of a large-scale hypertextual Web search engine // Comp. Networks and ISDN Systems. 1998. V. 30, iss. 1. P. 107–117.

16. Kleinberg J. M. Authoritative sources in a hyperlinked environment // J. of the ACM. 1999. V. 46, iss. 5. P. 604—632.

17. Маршакова И. В. Система связей между документами, построенная на основе ссылок: по данным Science Citation Index // НТИ. Сер. 2. 1973. № 6. С. 3–8.

18. Small H. Co-citation in the scientific literature: A new measure of the relationship between two documents // J. Amer. Soc. Inform. Sci. 1973. V. 24, iss. 4. P. 265–269.

19. Kessler M. M. Bibliographic coupling between scientific papers // Amer. Documentation. 1963. V. 14, iss. 1. P. 10–25.

20. Krauze T. K., McGinnis R. A matrix analysis of scientific specialties and careers in science // Scientometrics. 1979. V. 1, iss. 5–6. P. 419–444.

21. Lempel R., Moran Sh. SALSA: the stochastic approach for link-structure analysis // ACM Trans. Inf. Syst. 2001. V. 19, iss. 2. P. 131–160.

22. Freeman L. C. A set of measures of centrality based upon betweenness // Sociometry. 1977. V. 40. P. 35–41.

23. Altman A., Tennenholtz M. Ranking systems: the PageRank axioms / Proc. of the 6th ACM conf. on Electronic commerce (ACM). 2005. P. 1–8.

24. Kitti M. Axioms for centrality scoring with principal eigenvectors. [Electron. resource]. http://www.ace-economics.fi/kuvat/dp79.pdf.

Щербакова Наталья Григорьевна — старш. науч. сотр. Института вычислительной математики и математической геофизики СО РАН; e-mail: nata@nsc.ru

Дата поступления — 06.07.2015