МЕРЫ ЦЕНТРАЛЬНОСТИ В СЕТЯХ

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

УДК 001.12+303.2

Представлены два способа классификации мер центральности акторов сети: с точки зрения конкурирующих гипотез об отношении между структурными свойствами сети и поведением акторов и с точки зрения особенностей вычисления меры, отражающих вовлеченность узла в структуру маршрутов в сети.

Ключевые слова: анализ комплексных сетей, меры центральности сетевых акторов.

Two ways of centrality measures classification are presented: from the point of view of the competing hypotheses about the relation between structural properties of a network and behavior of actors, and according to the features of their calculation that assess a node's involvement in the walk structure of a network.

Keywords: complex networks analysis, node centrality measures.

Введение. Анализ комплексных сетей (*complex networks*), будь то сеть цитирования, социальная сеть и т. п., включает изучение топологических свойств сетей, рассматриваемых отвлеченно от их физической природы, но определяющих функционирование сетей. В этом случае сеть представляется в виде графа, вершины которого соответствуют акторам, а ребра — связям между ними. Активное изучение вопросов влияния топологии сети, отражающей взаимодействия акторов в социальных сетях на принятие групповых решений, началось с работ [1, 2], в которых была построена математическая модель связности акторов и предложены методы выявления коммуникационных потоков между ними. Также были определены понятия центральных и периферийных областей и сделано предположение, что между центральностью и влиянием в группе имеется связь. Эти работы дали толчок к проведению экспериментов, подтвердивших, что центральность, определяющая роль актора, является важным структурным атрибутом комплексных сетей. Мера центральности призвана ответить на вопрос: что характеризует важного актора? Ответ дается в терминах вещественной функции на вершинах графа, значения которой должны обеспечить ранжирование, выявляющее наиболее важные вершины. Однако единая концепция меры центральности и определение процедуры ее измерения отсутствовала, что инициировало изобретение множества мер центральности, часть которых слабо связана с интуитивным пониманием сущности. Рассмотрению основных интуитивных концепций, положенных в основу определений мер центральности, посвящена работа [3]. В ней классификация мер центральности актора отражает потенциал влияния на процессы, протекающие в сети. Однако при определении мер центральности делаются также неявные предположения о манере, в которой трафик транслируется по сети. Во внимание принимаются тип трафика (деньги, почта, инфекция и т. д.), механика обмена (копирование, передача) и выбор маршрута [4]. При графо-теоретическом подходе к центральности характеристикой ее мер являются структура рассматриваемых маршрутов и предполагаемое расположение актора на маршруте [5].

В данной работе рассматриваются основные меры центральности и их категоризация без учета различия между сетью и представляющим ее графом. Используемые термины приведены в *Приложении*.

1. Интуитивные предпосылки мер структурной центральности. Классификация мер центральности с точки зрения выявления определенного структурного свойства, связанного с интуитивной концепцией, предложена в работе [3]. Ряд определений центральности сводится к трем базовым, причем все три достигают максимума для центра сети, имеющей форму звезды или колеса.

1.1. *Центральность вершины как отражение потенциальной активности.* Простейшей, интуитивно понятной является концепция, в основе которой лежит предположение, что вершина с большим значением степени является потенциально активной. Идея использования степени вершины как индекса центральности, высказанная в работе [6], позднее была исследована в ряде работ, но наиболее простая, не зависящая от особенностей приложения мера (см., например, [7], [8]) приведена в работе [9]. Центральность по степени (*degree centrality*) C_D вершины i определяется как количество смежных с i вершин:

$$C_D(i) = \sum_{j=1}^{n} a(i,j), \tag{1}$$

где $a(i,j) = 1$ тогда и только тогда, когда вершины соединены ребром (здесь и далее n — количество вершин графа). Для графа без кратных ребер и петель $C_D(i)$ — это степень вершины i ($\deg(i)$). В общем случае $C_D(i)$ можно определить в терминах матрицы смежности $A = (a_{ij})$ графа как сумму по строке. В терминах матриц выражение (1) можно представить в виде $C_D = A\mathbf{1}$, где $\mathbf{1}$ — вектор, все элементы которого равны единице. В ориентированном графе различают входящую и исходящую C_D. В работе [10] предложен вариант определения, не зависящий от размеров сети, — это относительная центральность по степени:

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

Определенные в данном пункте меры можно назвать мерами, подобными C_D. Поскольку каждое ребро — это маршрут единичной длины, можно воспринимать центральность по степени как подсчет таких путей, начинающихся (оканчивающихся) в вершине [5]. Это специальный случай меры, предложенной в работе [11], названной *k-path centrality*, суть которой состоит в подсчете количества путей длиной k, выходящих из вершины. При $k = 1$ это центральность по степени, когда $k = n-1$ (максимальное значение) — количество всех путей любой длины, выходящих из вершины. Центральность может быть вычислена на основе сумм по строке матрицы W, w_{ij} — количество путей длиной меньше или равной k от вершины i до вершины j. Таким образом, $C_{k-path} = W\mathbf{1}$.

Мера, основанная на вычислении количества путей длиной k, не содержащих одинаковых ребер (ребро-независимых путей), начинающихся (оканчивающихся) в вершине (*edge-disjoint k centrality*), является вариацией *k-path centrality.* В работе [12] показано, что количество путей с непересекающимися ребрами между двумя вершинами равно максимуму потока, протекающего между ними, и доказана теорема, утверждающая, что количество путей с непересекающимися ребрами, связывающих две вершины, равно минимальному количеству ребер, которые нужно удалить, чтобы сделать вершины несвязанными, т. е. изолировать вершину. Также определены меры, основанные на подсчете количества путей длиной не более k, не содержащих пересекающихся вершин, кроме оконечных (вершино-независимых путей) — *vertex-disjoint k centrality*. Множество таких путей является подмножеством множества путей с непересекающимися ребрами. В работе [13] оценивается социальная близость (*social proximity*) — (нормализованный) счетчик всех путей с непересекающимися вершинами длиной два и меньше, связывающих вершины. В работе [14] определяется мера влияния *GPI*-индекса (*General Power Index*), для вычисления которого количество путей с непересекающимися вершинами четной длины, выходящих из вершины, вычитается из количества путей нечетной длины, выходящих из этой же вершины. Меры „достижимости“ (*reachability*), базирующиеся на подсчете количества вершин, которые можно достичь из заданной вершины за определенное число шагов (*vertex-disjoint k centrality*), приведены в работе [15].

Теперь рассмотрим так называемые „меры влияния“ (*eigenvector-like*), основанные на вычислении маршрутов, они вычисляются рекурсивно и косвенным образом свидетельствуют о центральности изучаемой вершины на основе данных о центральности соседних вершин. Примером такой меры является „центральность по Кацу“ [16].

Определим матрицу W, элементы которой вычисляются следующим образом:

$$w_{ij} = ba_{ij} + b^2(a^2)_{ij} + \cdots + b^k(a^k)_{ij} + \cdots = \sum_{k=1}^{n} b^k(a^k)_{ij}. \tag{2}$$

Центральностью вершины i называется величина $c_{KATZ}(i) = \sum\limits_{j=1}^{n} w_{ij}$. Матрица W имеет вид:

$$W = bA + b^2A^2 + b^3A^3 + \cdots = (I - bA)^{-1} - I,$$

где I — единичная матрица. Мера Каца (2) в матричном представлении выглядит так: $C_{KATZ} = W\mathbf{1}$, где $\mathbf{1}$ — вектор из единиц. Таким образом, если элементы матрицы смежности равны нулю или единице, то определяется взвешенный счетчик количества маршрутов, начинающихся или оканчивающихся в данной вершине. Маршруты взвешиваются величиной, обратной длине, так что длинные, в высшей степени непрямые маршруты стоят „дешево“, а короткие, прямые — „дорого“. Степень, до которой весы расходятся с длиной, контролируется параметром затухания b. Сходимость последовательности гарантируется только в том случае, когда $b < 1/\lambda$, где λ — максимальное собственное значение матрицы A. В работе [17] предложена мера, названная *total effects centrality* (*TEC*), подобная мере по Кацу. Центральность определяется с помощью матрицы

$$W = \left(I + bA + b^2A^2 + b^3A^3 + \cdots\right)(1 - b) = (I - bA)^{-1}(1 - b).$$

$$c_{TEC}(j) = \sum_i \frac{w_{ij}}{n-1}.$$

Несмотря на похожесть определений, в рамках модели, представленной в работе [17], меры разнятся.

Мера „центральность по Боначичу“ [18] сформулирована в более общих терминах, чем мера Каца. Определяется множество мер, зависящих от параметров α и β:

$$c_i(\alpha,\beta) = \sum_j (\alpha + \beta c_j) a_{ij}, \tag{3}$$

$$c\,(\alpha,\beta) = \alpha\,(I - \beta A)^{-1} A\mathbf{1}. \tag{4}$$

Параметр β определяет степень, до которой статус актора зависит от статуса тех, с кем он связан. При некоторых обстоятельствах β можно рассматривать как вероятность, а $c(\alpha,\beta)$ — как множество маршрутов, используемых акторами ($\beta < 1/\lambda$), т. е. перейти к формуле, подобной формуле Каца:

$$w_{ij} = \alpha\left(a_{ij} + b(a^2)_{ij} + b^2(a^3)_{ij} + \cdots\right) = \alpha \sum_{k=0}^{\infty} b^k (a^{k+1})_{ij}. \tag{5}$$

$$c_i = \sum_i w_{ij}.$$

В терминах матриц систему уравнений (5) можно представить в виде: $C_{BONACICH} = W\mathbf{1}$. Если $b = 0$, то формула определяет центральность по степени; если $b = 1/\lambda$, то это центральность собственного вектора [19]. Меры Каца и Боначича коррелируются только если b положительно. Вариант отрицательного b используется, например, при анализе обменных сетей (*exchange networks*) [20, 14]. Применение меры центральности собственного вектора к определенным типам данных можно найти в работах [21] (влияние (*power*)) и [22] (престиж (*prestige*)).

В определении меры по Хюббелю [23] рекурсивный способ вычисления, опирающийся на интуитивное представление о том, что центральность актора должна быть функцией от центральности с ним связанных акторов [19], заложен явно. Мера может быть представлена в виде:

$$C_{HUB} = XC_{HUB} + \mathbf{e},$$

где X — преобразованная матрица смежности, $\mathbf{e}$ — взвешивающий вектор. Если элементы вектора $\mathbf{e}$ — степени вершин, то это — центральность по Хойде [24].

Как показано в работе [25], собственный вектор матрицы смежности A является пределом меры Каца, когда b приближается к значению $1/\lambda$ из уравнения

$$\mathbf{v} = \lambda^{-1} A\mathbf{v}. \tag{6}$$

Собственный вектор $\mathbf{v}$ (предпочтительно главный) предложено рассматривать как „меру центральности собственного вектора“, подытоживающую определения, приведенные в [16, 23, 24]. Взвешивающий вектор добавляется в целях достижения сходимости процесса вычисления.

Особое место занимает алгоритм ссылочного ранжирования *PageRank*, предложенный в работе [26], так как он возобновил интерес к так называемому „спектральному ранжированию“ [27], основанному на вычислении собственного вектора преобразованной матрицы смежности. Рассматривается матрица, элемент которой a_{ij} равен единице, если со страницы i есть указатель на страницу j, и нулю в противном случае. Важность веб-страницы i определяется следующим образом:

$$x_i = \alpha \sum_j a_{ij} \frac{x_j}{k_j^{out}} + \beta.$$

Здесь α и β — константы, k_j^{out} — количество дуг, исходящих из вершины j. Если у вершины j нет исходящих дуг, то k_j^{out} приравнивается к единице, чтобы избежать деления на ноль. Уравнение для вектора центральности можно представить в виде

$$\mathbf{x} = \alpha A D^{-1} \mathbf{x} + \beta \mathbf{1},$$

где D — диагональная матрица с элементами $d_{ii} = \max(k_i^{out},1)$, а $\mathbf{1}$ — вектор размерности n, состоящий из единиц. Выражая $\mathbf{x}$, получим:

$$\mathbf{x} = \beta(I - \alpha A D^{-1})^{-1} \mathbf{1}.$$

Значение α должно быть меньше, чем значение, обратное максимальному собственному значению матрицы AD^{-1}. Сходство с мерой по Хюбелю очевидно.

Для оценки научных вкладов журналов определены две меры: *EigenFactor* и *ArticleInfluence* [28], вычисление которых основано на элементах главного собственного вектора преобразованной матрицы цитирования, представляющей взвешенный ориентированный граф.

1.2. *Центральность вершины как возможность управления передачей.* Другой взгляд на центральность вершины основывается на частоте прохождения через нее кратчайших (геодезических) путей между всеми парами вершин. Считается, что соответствующий актор может влиять на группу акторов, поддерживая, придерживая или разрушая процесс передачи информации. В любом случае он имеет потенциал для осуществления подобных действий. Особое внимание такому подходу уделено в работе [6], но формальное определение меры „центральность по посредничеству“ (*betweenness centrality*) дано в работах [29] (*напряжение* (*rush*)) и [3], в которых оно сформулировано для случая связных неориентированных графов. Однако „центральность по посредничеству“ применима и для ориентированных графов.

Пусть g_{ij} — количество кратчайших путей от вершины графа i до вершины j, а $g_{ij}(k)$ — количество кратчайших путей от i до j, проходящих через k. Тогда мера центральности $C_B(k)$ для вершины k определяется следующим образом:

$$C_B(k) = \sum_i \sum_j \frac{g_{ij}(k)}{g_{ij}} \quad (i \neq j \neq k)\,. \tag{7}$$

Если количество геодезических путей между i и j равно нулю, то принято считать, что $g_{ij}(k)/g_{ij} = 0$. Величина $g_{ij}(k)/g_{ij}$ определяет долю кратчайших путей между вершинами i и j, проходящих через k, ее можно интерпретировать как вероятность того, что случайно выбранный геодезический путь между i и j пройдет через k. Мера C_B, также как и мера C_D, зависит от размеров сети, поэтому значения распределяются в большом диапазоне, что приводит к необходимости нормализации. Одним из естественных способов нормализации является деление значения центральности на количество пар вершин n^2, т. е. „относительная центральность по посредничеству“ определяется как

$$C'_B(k) = \frac{1}{n^2} \sum_i \sum_j \frac{g_{ij}(k)}{g_{ij}} \ \ (i \neq j \neq k)\,.$$

При этом значения центральности будут находиться в диапазоне от нуля до единицы [30].

Варианты определения „центральности по посредничеству“ приведены в работе [4], в которой вместо геодезических путей рассматриваются все пути, все цепи и все маршруты (взвешенные обратно пропорционально длине). Аналитические формулы для их вычисления не приводятся, оценки выполнены на моделях. В работе [31] представлены аналитические формулы для модели случайного блуждания по маршрутам (*random-walk betweenness* (*RWB*)). В этом случае центральность также определяется формулой (5), но при этом подсчитывается отношение количества раз, когда случайно выбранный путь в среднем проходит через вершину i, к множеству повторений блуждания. Посредничество случайного блуждания определяется исходя из предположения, что путь продвижения трафика до пункта назначения выбирается случайным образом. Посредничество согласно кратчайшему пути несет полностью противоположный смысл. Тем не менее, в работе [31] показано, что на практике эти две меры дают сходные результаты.

„Потоковое посредничество“ (*flow betweenness*), предложенное в работе [32], базируется на понятии максимального потока. Представим, что по каждому ребру сети можно передать некоторый объем трафика, такой что за единицу времени общий объем переданного трафика между вершинами i и j будет максимальным. Для передачи одновременно можно использовать различные пути, пропускающие разные объемы трафика от i к j. Сформулируем вопрос: „Какая доля трафика проходит через вершину k, когда между вершинами i и j транслируется максимальный поток?“. Поскольку максимальный поток от i к j равен количеству ребро-независимых путей между i и j [30], то можно считать, что это доля независимых путей между i и j, проходящих через вершину k. Оба типа центральности (кратчайших путей и потоковая) отражают „оптимальные“ характеристики передающей среды, а именно прохождение по кратчайшим путям или передачу максимального объема. Таким образом, с помощью этих мер можно оценивать потенциал вершины по разрушению потока трафика путем прекращения ретрансляции.

Две меры, отражающие разные взгляды на связность узлов, приведены в работе [33]. Если определять связность в терминах достижимости, то мера „центральность по посредничеству“ является показателем фрагментации, возникающей после удаления узла. Мера F, названная „фрагментацией“ (*fragmentation*), — это пропорция несвязанных узлов сети, появляющихся в результате удаления узла. Чем больше значение, тем более важным является узел. Мера D_F (*distance-weighted fragmentation*), отражающая взгляд на связность в терминах дистанции, — это средняя обратная дистанция между узлами после удаления данного узла. Мера равна единице, если все узлы находятся на расстоянии единицы друг от друга, и равна нулю, если все узлы изолированы. Промежуточные значения отражают степень, до которой присутствие узла имеет тенденцию сокращать дистанции в сети.

1.3. *Центральность вершины как отражение ее независимости или эффективности.* Третья интуитивная концепция центральности вершины также базируется на идее контроля коммуникаций. Актор считается центральным до той степени, до которой он может избежать контроля со стороны других. Так, в работе [34] сделано предположение, что центральную позицию занимает актор, наиболее независимый от посредников при передаче информации. Независимость вершины графа определяется как „близость“ ко всем другим вершинам. В работе [1] приведен другой довод: сообщение, отправленное из наиболее центральной точки, дойдет до всех остальных за минимальное время (эффективность). Простой и естественный вариант такой меры представлен в работе [35], где центральность измеряется путем суммирования геодезических расстояний от вершины до всех остальных:

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

При таком определении вершина будет иметь максимальную центральность, если она наиболее удалена от остальных, так что эту меру можно считать мерой удаленности. Сумма дистанций имеет естественную интерпретацию в случае, если граф связный. В несвязном графе каждая вершина имеет бесконечную дистанцию хотя бы от одной точки, так что сумма превращается в бесконечность. Другие определения, являющиеся функциями от этой суммы [36–38], зачастую не имеют прозрачной интерпретации [10].

Технически меру удаленности можно преобразовать в меру близости, рассматривая величину, обратную значению $C_C(k)^{-1}$, умноженную на $n-1$ [39]:

$$C'_C(k) = \frac{n-1}{\sum\limits_j d_{kj}}.$$

Эта величина может трактоваться как обратное к средней дистанции, а так как $n-1$ — это минимальная сумма дистанций (вершина является смежной со всеми остальными), то и как обратное к доле, на которую вершина превосходит минимальное расстояние до остальных вершин.

Заметим, что значения имеют небольшую тенденцию динамического роста от наименьшего к наибольшему. Геодезическая дистанция между вершинами для большинства сетей имеет тенденцию принимать небольшое значение и расти только логарифмически по мере увеличения размера сети. Это означает, что разница между минимальной дистанцией, равной единице, и максимальной, порядка $\log n$, сама имеет порядок $\log n$, т. е. мала [30].

Другие варианты близости получаются путем варьирования определения матрицы дистанций между вершинами. Имеет смысл рассматривать все пути между вершинами i и j, а затем брать медиану или среднюю длину пути. Такой подход реализован в работе [17] при определении меры *IEC* (*Immediate Effects Centrality*), которая вычисляется как величина, обратная средней длине путей от вершины до всех остальных вершин. Проблема состоит в том, что эти усредняемые пути в общем случае не являются различными. Вопросы присвоения путям весов, оптимальных согласно разумному критерию, рассматриваются, например, в работах [40, 41]. Так в работе [41] дистанция между вершинами i и j преобразуется в „информационную близость“ взятием обратного значения $I_{ij} = 1/d_{ij}$, а центральность определяется как среднее гармоническое для ряда матрицы близости. Дана интерпретация матрицы близости с точки зрения теории информации, и мера названа „информационная центральность“ (*information centrality* (*IC*)):

$$IC(i) = \left[\frac{1}{n}\sum_j \frac{1}{I_{ij}}\right]^{-1}.$$

2. Типы мер с точки зрения вовлеченности в структуру сетевых маршрутов. Рассмотрим иную классификацию мер центральности [5] на примере ранее определенных.

Таблица

Перекрестная классификация мер центральности

Свойства маршрута	Позиция на маршруте	
	Радиальные меры	Медиальные меры
Объем	*Центральность по степени* [9, 10]; *k-path* [11]; *центральность собственного вектора* [25]; *центральность по Кацу* [16], *центральность по Хюббелю* [23], *центральность по Хойде* [24]; *GPI*-индекс [14]; *TEC* [17]; *влияние* [21]; *престиж* [22]; *центральность по Боначичу* [18]; *PageRank* [26]; *EigenFactor* [28]	*Центральность по посредничеству* [3]; *напряжение* [29]; *RWB* [31]; *потоковое посредничество* [32]; *MEC* [17]
Длина	*Центральность по близости* [35]; *IEC* [17]; *информационная центральность* [41]	D_F [33]

Заметим, что все меры центральности оценивают положение узла при блуждании по сетевым маршрутам. Варианты мер, подобных мере по степени, отличаются только ограничениями на то, какие маршруты подсчитываются: кратчайшие пути, все пути определенной длины, ребро-независимые, вершино-независимые. Таким образом, можно различать меры по типу маршрута (*walk type*). Кроме того, эти меры характеризуют длину маршрутов или их количество. На этом основании их можно назвать мерами, измеряющими длину (*length measures*) или объем (*volume measures*). Различие между мерами, измеряющими объем, и мерами, измеряющими длину, образует другое классификационное измерение, которое названо свойством маршрута (*walk property*). Меры, учитывающие маршруты, которые начинаются или заканчиваются в данной вершине, отнесем к категории радиальных (*radial*). В отличие от этого меры, подобные центральности по посредничеству, учитывающие количество маршрутов, проходящих через какую-либо вершину, отнесем к категории медиальных (*medial*). Исключение составляет мера D_F, определенная в терминах дистанции. Различие между радиальными и медиальными мерами образует третье классификационное измерение — позицию на маршруте (*walk position*). Рассмотренные меры основаны на суммах по строке или столбцу матрицы смежности или преобразованной матрицы смежности графа. Обычно это простая сумма или среднее. Однако используются и другие способы суммирования, такие как взвешенные средние, медианы, моды, минимумы или максимумы. Важным примером взвешенной суммы являются собственные векторы. Классификационное измерение, учитывающее способ суммирования, названо „тип суммирования“ (*summary type*). Рассуждения подытожены в таблице, для простоты включающей только два измерения: позицию на маршруте и свойства маршрута.

Такая классификация делит меры на группы, разумнее сравнивать между собой меры внутри группы, чем члены разных групп. Меры в разных группах различаются фундаментально, однако их можно рассматривать как дополняющие друг друга. Меры внутри группы схожи по ключевым признакам, так что их можно рассматривать как альтернативные при ответе на вопрос о том, какую меру следует применить.

Меры радиальной центральности имеют разумную сетевую интерпретацию применительно к сети, в которой парная связность однородна. В этом случае радиальная центральность служит эффективным мерилом принадлежности к ядру [42, 43]. В работе [19] отмечено, что если сеть содержит более одной компоненты (максимальное множество взаимосвязанных узлов), то центральность собственного вектора узлов, находящихся вне наибольшей компоненты, будет равна нулю. Медиальные меры приписывают особо большие значения узлам, являющимся мостами в сети. Однако трудно интерпретировать меру медиальной центральности без знания сетевой топологии.

3. Центральность сети. Перейдем к понятию центральности сети (графа). Одни авторы рассматривают центральность с точки зрения компактности графа, соответствующие меры основываются на дистанциях между вершинами. Считается, что граф централен до той степени, до которой точки близки друг к другу [2, 39]. Альтернативный подход — рассматривать центральность графа как степень отличия значения меры центральности вершины с наибольшим значением от значений для других вершин [34, 9, 10]. В такой интерпретации центральность графа должна отражать тенденцию определенной вершины быть наиболее центральной, т. е. централизацию графа.

В идеальном случае индексы централизации графа, независимо от того, на какой мере центральности вершины они основаны, должны обладать следующими свойствами [10]: 1) показывать, до какой степени центральность самой центральной вершины превосходит центральность остальных вершин; 2) выражаться отношением этого превосходства к максимально возможному значению для графа, содержащего рассматриваемое количество вершин. Т. е., если $C_G(v_i)$ — центральность вершины v_i графа $G(V,E)$, а $C_G(v^*)$ — максимальное значение центральности вершины, то центральность графа определяется как

$$C_G = \frac{\sum\limits_{i=1}^{n}[C_G(v^*) - C_G(v_i)]}{\max\sum\limits_{i=1}^{n}[C_G(v^*) - C_G(v_i)]}.$$

При этом $C_G = 0$ тогда и только тогда, когда центральности всех вершин равны; $C_G = 1$ тогда и только тогда, когда v^* полностью доминирует над остальными [10].

Например, „центральность по посредничеству" графа в работе [3] определяется следующим образом:

$$C_B = \frac{\sum\limits_{i=1}^{n}(C'_B(v^*) - C'_B(v_i))}{n-1},$$

где v^* — вершина, получившая наибольшее значение центральности.

Фундаментальное различие между радиальными и медиальными мерами центральности графа выявлено в работе [44]. В ней рассмотрены все возможные связные графы, состоящие из 4, 5, 6, 7 и 8 вершин, и вычислены три меры центральности графов, определенные согласно [10]. Сделан вывод, что между центральностью по степени и центральностью по близости существует линейная зависимость, а центральность по посредничеству имеет существенное отличие. Кроме того, замечено, что граф, меры центральности которого могут быть упорядочены как $C_B > C_C > C_D$ (а), имеет тенденцию быть разделенным на подгруппы, соединенные между собой ребром или общей центральной точкой. Этот тип графа может быть охарактеризован как децентрализованная сеть. С другой стороны, если выполняется соотношение $C_C > C_D > C_B$

(б), то граф представляет собой сеть, состоящую из одного кластера, в который входит большинство узлов. Исключение составляют только графы-цепи. Количество ребер является важным параметром для сравнения мер. Если количество ребер мало, центральность по посредничеству графа имеет в среднем большое значение, а центральность по близости — маленькое; с возрастанием количества ребер порядок (а) сменяется на (б).

Заключение. Меры центральности рассмотрены с точки зрения концептуальной и графо-теоретической основы. Наличие большого количества определений меры центральности связано с неоднозначностью самого понятия центральности. В работе [10] предложены три интуитивных категории, к которым относится большинство определений центральности вершины: вершина с большим значением степени является потенциально активной; вершина, через которую проходит много маршрутов между другими вершинами, контролирует связь; вершина, которая может избежать контроля со стороны других, является независимой. Три типа центральности выражают три взгляда на то, как центральность может влиять на процессы, протекающие внутри группы акторов.

Графо-теоретическая концепция предлагает взгляд на центральность с точки зрения вовлеченности узла в структуру маршрутов сети. Меры различаются по четырем ключевым измерениям: позиция на маршруте, тип маршрута, свойства маршрута и выбор способа суммирования в матрице смежности. В одних случаях нужно оценить риск своевременной доставки чего-либо, тогда больше подходят меры, учитывающие длину маршрута, если рассматривается целостность объема, то меры объема. Разница между радиальными и медиальными мерами более существенна. Они вместе оценивают вовлеченность узла в коммуникации, дополняя друг друга [17].

Вне рассмотрения остались алгоритмы вычисления мер центральности, которые зачастую имеют большую вычислительную сложность. Так, меры, основанные на кратчайших путях, в общем случае имеют сложность $O(n^3)$, что препятствует их использованию в больших реальных сетях. Поэтому предлагается решение проблемы путем разработки алгоритмов аппроксимации [45] и даже разработки новых мер центральности, в определение которых заложены аппроксимация и возможность параллельного вычисления [46].

Приложение (по материалам [47, 48]).

П.1. *Граф* — это упорядоченная пара $G = (V, E)$, где V — непустое множество вершин, а E — множество пар вида $e = (u, v)$, u, $v \in V$, называемых в соответствии с геометрическим представлением ребрами, а вершины u и v — концами ребра. Вершины ребра (u, v) называются смежными, а ребро e и вершина u (также как и v) — *инцидентными.* Если пары упорядочены, то граф называется *ориентированным*, в противном случае — *неориентированным.* Будем рассматривать конечные графы, когда V — конечное множество мощности n. Граф называется *взвешенным*, если задана весовая функция $f : E \to \mathbb{R}$, сопоставляющая с каждым ребром его вес.

П.2. *Матрица смежности* графа $G = (V, E)$ с конечным числом вершин n (пронумерованных числами от 1 до n) — это квадратная матрица A размера n, в которой значение элемента a_{ij} равно числу ребер из i-й вершины графа в j-ю. Если граф не имеет кратных ребер, то $a_{ij} = 1$, если ребро $(i, j) \in E$.

П.3. *Единичная матрица* — это квадратная матрица, элементы главной диагонали которой равны единице, остальные — нулю.

П.4. *Маршрут* — это конечная последовательность ребер, так что любые два последовательных ребра имеют общую концевую точку. *Цепь* — маршрут, в котором каждое ребро встречается не более одного раза. *Простой цепью* называют маршрут, в котором каждая вершина встречается не более одного раза. Простую направленную цепь будем называть путем.

П.5. *Цикл* — это замкнутая цепь. *Простой цикл* — цикл, вершины которого не повторяются, их количество больше 3. *Петля* — ребро, у которого оба конца совпадают, т. е., $e = (u, u)$. Два ребра называются *кратными*, если множества их концевых вершин совпадают. Мы будем рассматривать графы без петель и кратных ребер.

П.6. Неориентированный граф называется *связным*, если для каждых двух вершин $u, v \in V$ существует путь из u в v. Ориентированный граф называется слабо связным, если связным является соответствующий ему неориентированный граф, и называется сильно связным, если для каждых двух вершин u, v существует ориентированный путь из u в v.

П.7. *Расстояние* (геодезическое) $d(u,\ v)$ между вершинами u и v в связном невзвешенном неориентированном графе — это длина наименьшей простой цепи между u и v. Вес пути во взвешенном графе — это сумма весов входящих в него ребер. *Кратчайшим путем* между вершинами u и v во взвешенном графе считается путь, имеющий минимальный вес, а длина пути равна весу. Если в графе вершины u и v несвязны, то считается, что $d(u,\ v) = \infty$.

П.8. *Степень* вершины $\deg(v)$ — это количество инцидентных с вершиной v ребер (петли считаются дважды). В случае ориентированных ребер различают входящую (*in-degree*) и исходящую (*out-degree*) степени (другое называние — полустепень захода и полустепень исхода).

П.9. *Граф-цепь (chain graph)* — это граф, не содержащий направленных циклов.

П.10. *Граф-звезда* (*star*) S_k — это дерево с одной внутренней вершиной и k листьями.

П.11. *Колесом (wheel)* W_n называется граф с n вершинами ($n \geq 4$), образованный соединением единственной вершины со всеми вершинами $(n-1)$-цикла.

Список литературы

1. Bavelas A. A mathematical model for group structure // Human Organization. 1948. V. 7. P. 16–30.

2. Bavelas A. Communication patterns in task-oriented groups // J. the Acoustical Soc. of Amer. 1950. V. 22. P. 271–288.

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

4. Borgatti S. P. Centrality and network flow // Soc. Networks. 2005. V. 27, P. 55–71.

5. Borgatti S. P., Everett M. G. A graph-theoretic perspective on centrality // Soc. Networks. 2006. V. 28. P. 466–484.

6. Shaw M. E. Group structure and the behavior of individuals in small groups // J. Psychology. 1954. V. 38. P. 139–149.

7. Rogers D. L. Sociometric analysis in interorganizational relations: application of theory and management // Rural Sociology. V. 39. P. 487–503.

8. Czepiel J. A. Word of mouth processes in the diffusion of major technological innovation // J. Marketing Res. V. 11. P. 172–180.

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

10. Freeman L. C. Centrality in social networks. Conceptual clarification // Soc. Networks. 1978/79. V. 1. P. 215–239.

11. Sade D. S. Sociometrics of macaca mulatta III: N-path centrality in grooming networks // Soc. Networks. 1989. V. 11. P. 273–292.

12. Ford L. R., Fulkerson D. R. Maximal flow through a network // Canadian J. Math. 1956. V. 8. P. 399–404.

13. Alba R., Kadushin C. The intersection of social circles // Sociol. Meth. and Res. 1976. V. 5. P. 77–102.

14. Markovsky B., Willer D., Patton T. Power relations in exchange networks // Amer. Sociol. Rev. 1988. V. 53. P. 220–236.

15. Higley J., Hoffman-Lange U., Kadushin C., Moore G. Elite integration in stable democracies: a reconsideration // Europ. Sociol. Rev. 1991. V. 7. P. 35–53.

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

17. Friedkin N. E. Theoretical foundations for centrality measures // Amer. J. Sociol. 1991. V. 96. P. 1478–1504.

18. Bonacich P. F. Power and centrality: A family of measures // Amer. J. Sociol. 1987. V. 92. P. 1170–1182.

19. Bonacich P. Factoring and weighting approaches to status scores and clique identification // J. Math. Sociol. 1972. V. 2. P. 113–120.

20. Cook K. S., Emerson R. M., Gillmore M. R., Yamagishi T. The distribution of power in exchange networks: theory and experimental results // Amer. J. Sociol. 1983. V. 89. P. 275–305.

21. Coleman J. S. Loss of power // Amer. Sociol. Rev. 1973. V. 38. P. 1–17.

22. Burt R. S. Toward a structural theory of action. N.Y.: Academic Press, 1982.

23. Hubbell C. H. An input-output approach to clique identification // Sociometry. 1965. V. 28. P. 377–399.

24. Hoede C. A new status score for actors in a social network // Twente University, dept. of applied mathematics (Internal report). 1978; [Electron. resource]. `http://doc.utwente.nl/64930/1/memo1877.pdf`.

25. Bonacich P. F. Simultaneous group and individual centralities // Soc. Networks. 1991. V. 13. P. 155–168.

26. Page L., Brin S., Motwani R., Winograd T. The pagerank citation ranking: Bringing order to the web // Technical report, Stanford Digital Library Technologies Project. 1998. [Electron. resource]. `http://ilpubs.stanford.edu:8090/422/`.

27. Vigna S. Spectral Ranking. [Electron. resource]. `http://vigna.dsi.unimi.it/papers.php`.

28. Bergstrom C. T., West J. D. Assessing citations with the eigenfactor metrics // Neurology. 2008. V. 71. P. 1850–1851.

29. Anthonisse J. M. The rush in a directed graph // Technical report BN 9/71. Amsterdam: Stiching Matematisch Centrum, 1971.

30. Newman M. E. J. Networks. An introduction. N.Y.: Oxford University Press, 2010.

31. Newman M. E. J. A measure of betweenness centrality based on random walks // Soc. Networks. 2005. V. 27. P. 39–54.

32. Freeman L. C., Borgatti S. P., White D.R. Centrality in valued graphs: A measure of betweenness based on network flow // Soc. Networks. 1991. V. 13. P. 141–154.

33. Borgatti S. P. Identifying sets of key players in a network // Computational, Mathematical and Organizational Theory. 2006. V. 12, iss. 1. P. 21–34.

34. Leavitt H. J. Some effects of certain communication patterns on group performance // J. Abnorm. and Soc. Psychology. 1951. V. 46. P. 38–50.

35. Sabidussi G. The centrality index of a graph // Psychometrika. 1966. V. 31. P. 581–603.

36. Høivik T., Gleditsch N. P. Structural parameters of graphs: a theoretical investigation. In quantitative sociology. N.Y.: Academic Press, 1975. P. 203–223.

37. Valente T. W., Foreman R. K. Integration and radiality: measuring the extent of an individual's connectedness and reachability in a network // Soc. Networks. 1998. V. 20. P. 89–105.

38. Burt R. S. Structure Reference Manual. Ver. 4.2. N.Y.: Center for the Social Sciences, Columbia University, 1991.

39. Beauchamp M. A. An improved index of centrality // Behav. Sci. 1965. V. 10. P. 161–163.

40. Nunnally J. C. Psychometric theory. N.Y.: McGraw Hill, 1967.

41. Stephenson K., Zelen M. Rethinking centrality: methods and examples // Soc. Networks. 1989. V. 11. P. 1–37.

42. Borgatti S. P., Everett M. G. Models of core/periphery structures // Soc. Networks. 1999. V. 21. P. 375–395.

43. Everett M. G., Borgatti S. P. Extending centrality. In Models and Methods in Social Network Analysis. Cambridge: Cambridge University Press. 2005.

44. Nakao K. Distribution of measures of centrality: enumerated distributions of freeman's graph centrality measures // Connections. 1990. V. 13, N. 3. P. 10–22.

45. Bader D. A., Kintali Sh., Madduri K., Mihail M. Approximating betweenness centrality // Workshop on algorithms and models for the web-graph, San Diego (USA), December 11–12. [Electron. resource]. `http://www.cc.gatech.edu/~bader/papers/ApproxBC.html`.

46. Kang U., Papadimitriou S., Sun J., Tong H. Centralities in large networks: algorithms and observations // Proc. of the Eleventh SIAM International Conference on Data Mining, Mesa (USA), Apr. 28–30, 2011. P. 119–130.

47. Оре О. Теория графов. М.: Наука. 1980.

48. Харари Ф. Теория графов. М.: Мир. 1973.

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

Дата поступления — 29.04.2015