

## Центральность распада в социальных графах и адаптированный алгоритм Флажолле-Мартена для ее расчета

Торопов Б.А.

**Аннотация**— В статье рассматривается модель центральности вершин в графе, которая является общим случаем центральности по степени и центральности по близости для графов, отражающих социальные сети. Эта форма центральности основана на степенной функции, предложенной Джексоном и Волински как центральность распада. Автор рассматривает возможность аппроксимированной оценки центральности распада для вершин графа на основе адаптированного алгоритма Флажолле-Мартена. По результатам расчета сравниваются полученные аппроксимированные значения центральности распада с реальными значениями центральности по близости. Методологию исследования составляют элементы теории графов, а также аппарат анализа социальных сетей, связанный с расчетом метрик центральности. Вычислительный эксперимент показал, что данная форма центральности подходит для оценки центральности по близости в социальных графах. А предложенный алгоритм подходит для аппроксимированной оценки значений центральности вершин графа, связанных с кратчайшими путями, таких как центральность распада и дает существенный выигрыш в скорости вычислений.

**Ключевые слова**— социальный граф, социальная сеть, центральность, алгоритм Флажолле-Мартена, аппроксимация.

### 1. ВВЕДЕНИЕ

Центральность вершины в графе [1] – ключевое понятие в анализе социальных сетей. К настоящему времени исследователями в данной области предложена масса метрик центральности, таких как, центральность по степени, центральность по близости, центральность по промежуточности, собственный вектор и многие другие, каждая из которых позволяет оценить важность вершины в графе по сравнению с другими вершинами. Будучи рассчитанной для каждой вершины социального графа, любая метрика центральности, например, из числа упомянутых выше, позволяет определить к наиболее значимых вершин.

Что собственно понимать под значимостью вершины и какая из метрик применима к решению той или иной аналитической задачи, зависит от постановки задачи, природы исследуемого графа, от того какие именно социальные объекты, процессы и явления принимаются за его вершины и ребра.

Так, например, в [2] авторами рассмотрен пример сегмента террористической сети, состоящей из спящих

ячеек, каждая из которых практически изолирована от остальных, а коммуникации между ячейками осуществляются посредством небольшого количества участников сети, обладающих высокой центральностью по промежуточности (их посреднические возможности велики по сравнению с другими участниками).

Для небольших сообществ, рассматриваемых герметично, в изоляции от внешнего мира, скорее всего, будет наиболее важна центральность по степени (числу прямых контактов с остальными участниками).

При решении некоторых практических задач «традиционные» метрики центральности (отнесем к ним степень, близость и промежуточность) показывают несостоятельность отразить реальные социальные сети и происходящие в них процессы, а наиболее подходящими оказываются более изощренные характеристики участников сети, такие как, например, собственный вектор. Как показано в работе [3] именно расчет собственного вектора участников сети позволил выявить среди жителей деревенских сообществ юго-запада Индии тех, которые в наибольшей степени способствуют внедрению в деревенскую жизнь новых финансовых инструментов.

Центральность по близости зачастую рассматривается как метрика, показывающая, насколько быстро информация распространится по сети, если инициатором распространения будет заданная вершина [4]. При продвижении идей, товаров или услуг в онлайн-новых социальных сетях, субъекты, стоящие за продвигаемой информацией, могут быть заинтересованы в определении тех пользователей, которые, будучи проводниками этой информации, в кратчайшие сроки обеспечат наибольший охват аудитории. Такие участники, скорее всего, должны обладать высокой близостью ко всем прочим участникам сети в смысле длин кратчайших путей. Хотя эта идея вполне очевидна, на сегодняшний день остается актуальным научно-практическим направлением поиск эффективных методов расчета метрик центральности, связанных с кратчайшими путями, масштабируемых для применения на действительно больших графах, существующих в реальном мире, таких как наиболее популярные онлайн-социальные сети.

### II. ВЫЧИСЛИТЕЛЬНАЯ СЛОЖНОСТЬ РАСЧЕТА ЦЕНТРАЛЬНОСТИ ПО СТЕПЕНИ И ЦЕНТРАЛЬНОСТИ ПО БЛИЗОСТИ

Далее будем использовать формальное представление графа следующего вида:

$$V = \{v_1, v_2 \dots v_n\}; g = \{0, 1\}^{V \times V}, \quad (1)$$

где:  $V$  – множество участников сети;

Статья получена 20 июля 2017.

Торопов Борис Андреевич, доцент кафедры информационных технологий Академии управления МВД России (эл. почта: torbor@mail.ru)

$g$  – матрица сопряженности.

Каждый элемент квадратной матрицы  $g$  определяет наличие связи между парой вершин графа  $vu$  и принимает значение 1, если связь между вершиной по строке  $v$  и вершиной по столбцу  $u$  есть, и 0, если связи нет.

Центральность по степени показывает число непосредственных связей вершины  $v$  с другими вершинами. Для получения значения центральности по степени необходимо просуммировать все единицы в строке, соответствующей вершине  $v$ , из матрицы сопряженности:

$$c_v^d = \sum_u g_{vu} \quad (2)$$

Алгоритм выполняется за одну итерацию. Для расчета степени для каждой вершины графа требуется «обход» алгоритмом  $N$  вершин.

Расчет близости значительно более трудоемок. Центральность по близости – это метрика, показывающая насколько вершина близка ко всем остальным вершинам графа в смысле длин кратчайших путей:

$$c_v^{cl} = \frac{N-1}{\sum_u l_{vu}}, \quad (3)$$

где  $l_{vu}$  – длина кратчайшего пути между  $v$  и  $u$ .

Исходя из этого прямой расчет центральности по близости предполагает определение кратчайших путей от анализируемой вершины до всех остальных вершин графа, иначе говоря, для расчета требуется решение задачи о кратчайшем пути в заданный пункт назначения (single-source shortest path (SSSP) problem) для данной вершины. Решение задачи алгоритмом поиска в ширину (breadth first search (BFS) algorithm) в наилучшем возможном случае повлечет вычислительную сложность  $O = (VE+V^2)$  операций, поскольку поиск кратчайшего пути между парой вершин, опять же в наихудшем случае, занимает  $O = (E+V)$ , где  $E$  – это число ребер графа. Алгоритм Дейкстры позволяет решить данную задачу за  $O = (E+V \times \log_2 V)$  [5]. Но если требуется найти кратчайшие пути между всеми возможными парами вершин, то сложность алгоритмов BFS и Дейкстры возрастает еще в  $V$  раз до  $O = (V^2 E + V^3)$  и  $O = (VE + V^2 \times \log_2 V)$ , соответственно. Кроме того, пространственная сложность этих алгоритмов также велика, для работы оба они требуют хранения в памяти компьютера существенного объема данных, связанных с уже «обойденными» вершинами, на каждой итерации выполнения алгоритма. Для больших графов это также зачастую мало приемлемо.

В настоящей работе рассматривается метод «быстрого» аппроксимированного расчета центральности, которая является обобщенным случаем центральности по степени и центральности по близости.

### III. ОБОБЩЕННАЯ ФОРМА ЦЕНТРАЛЬНОСТИ ПО СТЕПЕНИ И ПО БЛИЗОСТИ НА ОСНОВЕ ЦЕНТРАЛЬНОСТИ РАСПАДА

Центральности по степени и по близости могут быть приведены к обобщенной форме, например, так, как это показано в работе Жао и др. [6], где задается  $g: \mathbb{R} \geq 0 \mapsto \mathbb{R} \geq 0$  – монотонно убывающая всюду

положительная функция от расстояния между вершинами  $r$ . Также обозначим количество вершин, находящихся от  $v$  на расстоянии не более  $r$  как  $N(v, r)$ , а число вершин, находящихся от  $v$  на расстоянии ровно  $r$  – как  $N_v(r)$ . Пользуясь этими обозначениями приведем обобщенную формулу центральности согласно Жао и др.:

$$c_v = \sum_r g(r) N_v(r) = \sum_r g(r) [N(v, r) - N(v, r-1)] \quad (4)$$

При  $r=1$  это выражение будет определять центральность по степени, в других случаях, когда  $r \in (1:d]$ , где  $d$  – диаметр графа, центральность будет принимать различные значения в зависимости от формы  $g(r)$ .

Рассмотрим случай, когда за  $g(r)$  принята степенная функция вида  $\delta^r$ ,  $\delta \in (0:1)$ . В этом случае получим метрику центральности, введенную Джексоном и Вольнски [7], и впоследствии обозначенную как центральность распада (decay centrality) [8].

$$c_v^{dec} = \sum_{u \neq v} \delta^{l(v,u)}, \quad (5)$$

где  $l(v,u)$  – длина кратчайшего пути между вершинами  $v$  и  $u$ .

Степенная форма  $g(r)$  позволяет для некоторых графов сделать обобщение между центральностью по степени и центральностью по близости. Действительно, при  $\delta \rightarrow 0$ , значимость непосредственных соседей вершины  $v$  будет много больше, чем вершин, находящихся от  $v$  на расстоянии 2 и дальше, то есть расчетные значения  $c_v^{dec}$  будут прямо коррелированы

со значениями центральности по степени  $c_v^d$ , тем в большей степени, чем ближе к нулю значение коэффициента распада  $\delta$ . В то же время при достаточно больших значениях  $\delta$  в районе  $0,7 - 0,9$  расчетные значения центральности  $c_v^{dec}$  будут коррелированы с

центральностью по близости  $c_v^{cl}$  для графов с небольшим  $d$  (или небольшим средним кратчайшим путем между вершинами, который не столь чувствителен к выбросам, как диаметр), поскольку в этом случае значимость соседей  $v$  с увеличением расстояния до них будет убывать похожим образом (рис.1). Конечно, в сильно разреженных графах центральность распада в отличие от центральности по близости будет придавать существенно большую значимость наименее удаленным от  $v$  вершинам, и сводить значимость большого количества дальних соседей к ничтожно малым величинам.

Вместе с тем, существует множество экспериментальных подтверждений тому, что в реальных социальных сетях, таких как непосредственные коммуникации между людьми или опосредованные – путем почтовых сообщений или онлайн, средний кратчайший путь невелик. Этот феномен известен сегодня как Теория шести рукопожатий (Six degrees of separation). В пионерской работе на данную тему, Милгрэм [9] описал результаты эксперимента, проведенного в США и показавшего, что письма, направленные без указания

точного адреса получателя, но с просьбой о дальнейшей пересылке тому, кто может его знать, находили своего адресата с числом пересылок от 2 до 10, а медианное значение кратчайшего пути равнялось 5 (для 25% писем

из общего числа отправленных, которые дошли до адресата).

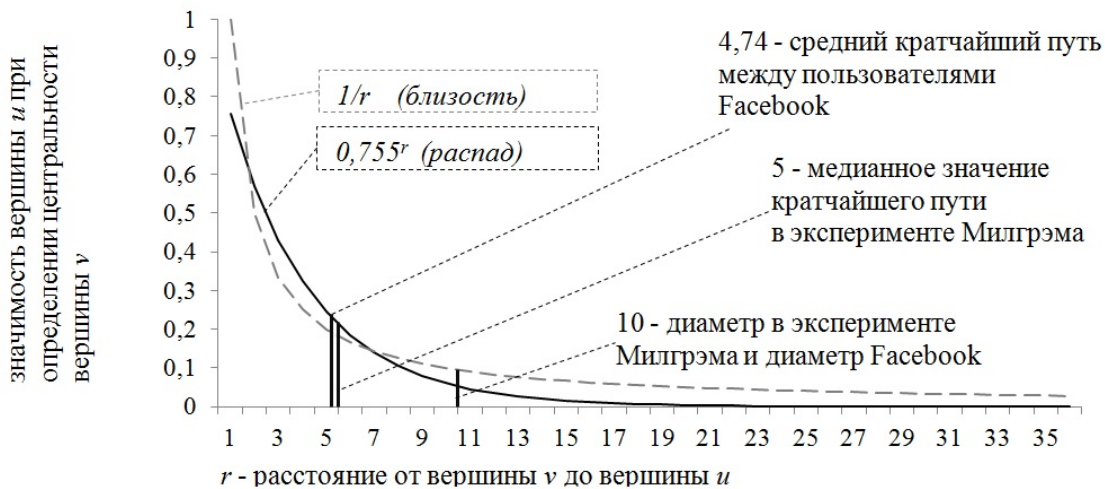


Рис.1. Значимость вершины  $u$ , удаленной от вершины  $v$  на расстояние  $r$ , в случае расчета центральности по близости и центральности распада

В конце прошлого века (1999) средний кратчайший путь между сайтами Всемирной паутины оценивался в 3,1 [10].

Средний кратчайший путь между пользователями онлайн-социальной сети Facebook на относительно современном этапе (2012 г., а по некоторым источникам [11] – это пик пользовательской активности в данной сети) составляет 4,74 [12]. Диаметр этой социальной сети оценивается в 10 [13] что совпадает с результатом, полученным Милгрэмом в 1967 г. для почтового сообщения в США.

Как видно из рис.1 на участке  $r \in [1:10]$  кривые, характеризующие значимость участников, удаленных от  $v$  на расстояние  $r$ , при расчете центральности по близости и центральности распада, ведут себя схожим образом. Это позволяет предложить обобщенную форму центральности для графов с небольшим средним кратчайшим путем на основе центральности распада Джексона. Пользуясь (4) и (5) запишем ее формальное представление:

$$c_v^{dec} = \sum_r \delta^r N_v(r) = \sum_r \delta^r [N(v,r) - N(v,r-1)] \quad (6)$$

Нормировать полученные значения центральности вершин можно по  $\delta(V-1)$  – максимально возможному значению данной метрики (случай, когда вершина  $v$  находится в непосредственном соседстве со всеми остальными  $V-1$  вершинами):

$$c_v^{dec} = \sum_r \delta^r [N(v,r) - N(v,r-1)] / \delta(V-1) \quad (7)$$

Предложенная форма центральности также не лишена недостатка, связанного с существенной временной и пространственной сложностью ее расчета при использовании алгоритмов BFS или Дейкстры.

Однако сегодня существует эффективный метод, позволяющий вычислить аппроксимированное значение  $N(v,r)$ .

#### IV. МЕТОДИКА РАСЧЕТА ЦЕНТРАЛЬНОСТИ РАСПАДА НА ОСНОВЕ АЛГОРИТМА ФЛАЖОЛЕ-МАРТЕНА

Алгоритм Флажолле-Мартена изначально предложен авторами для аппроксимированного подсчета уникальных записей в базах данных [14]. В ряде работ [4, 6, 15] показано, что он адаптируется для подсчета числа вершин в графе, которые удалены от заданной вершины на расстояние не более  $r$ . Вычислительная сложность алгоритма составляет  $O = (dE)$ , что значительно меньше, например,  $O = (VE + V^2 \times \log_2 V)$ , предлагаемых алгоритмом Дейкстры, поскольку число ребер графа в наихудшем случае равняется  $V^2$ , при  $d=1$ .

Согласно алгоритму Флажолле-Мартена каждой вершине графа присваивается кодовая строка  $M(v,0)$ , формируемая следующим образом. С вероятностью  $1/2$  нулевому биту кодовой последовательности задается значение 1, с вероятностью  $1/4$  – единичное значение присваивается первому биту,  $1/8$  – второму, и т.д. Каждому  $i$ -му биту кода с вероятностью  $p = 2^{-(i+1)}$  присваивается значение 1, и с вероятностью  $q = 1-p$  – присваивается значение 0.

Длина кодовой строки у всех вершин фиксированная, и составляет не менее  $\log_2 N$  бит, что позволит рассчитать случай нахождения предельно возможного числа вершин на одинаковом удалении от  $v$ , так как бинарная строка длины  $\log_2 N$  позволяет закодировать ровно  $N-1$  десятичных чисел.

На каждой итерации  $r$  выполнения алгоритма кодовая строка каждой  $M(v,r)$  задается равной логическому ИЛИ ( $\oplus$ ) между собственным значением на предыдущей итерации и значениями кодовых строк вершин – непосредственных соседей также на предыдущей итерации.

Дальнейший расчет основывается на следующем положении: для того, чтобы получить в кодовой комбинации  $M(v,r)$  некоторое количество  $b$  единичных разрядов перед первым нулевым, необходимо при

помощи логического ИЛИ «обойти» такое количество вершин графа, которое пропорционально  $2^b$ . Это позволяет дать оценку тому количеству вершин графа, которое находится на расстоянии кратчайшего пути не более  $r$  от заданной вершины  $v$ .

Таким образом, для вершины  $v$  определяется значение  $b$  – порядок первого разряда в  $M(v,0)$ , значение которого равно нулю. Затем определяется  $\hat{N}(v,r)$  оценочное количество вершин находящихся от  $v$  на длине кратчайшего пути не более  $r$ , где  $r$  – это номер текущей итерации алгоритма.  $\hat{N}(v,r)$  пропорционально  $2^b$  и рассчитывается следующим образом:

$$\hat{N}_v(r,v) = \frac{2^b}{0,77359}, \quad (8)$$

где 0,77359 – поправочный коэффициент, согласно [14].

Псевдокод данного алгоритма, адаптированный для расчета центральности распада по выражению (7) представлен на рисунке 2.

Предложенная обобщенная форма центральности на основе центральности распада применима, например, для моделирования распространения сообщений в онлайн-социальных сетях. Она отражает особенность этих сетей, предполагающую опосредованное влияние одного пользователя на других, которые не находятся с ним в формально зафиксированных отношениях (в друзьях, фоловерах, подписчиках и т.п., в зависимости от обозначения принятого в конкретной онлайн-сети) посредством участия в комментировании на страницах общих друзей, друзей друзей и т.д. Сила такого влияния с возрастанием длины пути между пользователями сети затухает согласно закону, который предложенная форма центральности отражает. Значения центральности могут быть рассчитаны на основе адаптированного алгоритма Флажолле-Мартена, как показано выше.

## V. ВЫЧИСЛИТЕЛЬНЫЙ ЭКСПЕРИМЕНТ

Расчеты центральности распада по алгоритму Флажолле-Мартену проводились на случайных  $G(n,p)$ -графах.  $G(n,p)$ -граф состоит из  $n$  вершин, между каждой возможной парой которых с вероятностью  $p$  есть ребро, согласно модели, впервые предложенной Эрдешем-Реньи в [16]. Здесь для обозначения количества вершин будем вместо  $n$  по-прежнему придерживаться  $V$ . Для расчета центральностей, из сгенерированного графа, состоящего из  $V = 500$  вершин, выделялся наибольший связный компонент, его величина обозначена как  $V_c$ , коэффициент кластеризации – как  $k_{clust}$ , длина среднего кратчайшего пути –  $\bar{l}$  и диаметр –  $d$ .

Для вершин в полученных графах были рассчитаны центральность по близости  $c_v^{cl}$  и центральность распада

$c_v^{dec}$  с коэффициентом распада  $\delta = 0,8$  на основе адаптированного алгоритма Флажолле-Мартена. Результаты показали наличие выраженной взаимозависимости между  $c_v^{dec}$  и  $c_v^{cl}$  (рис.3).

Наиболее сильная взаимозависимость наблюдается в графах, где средняя длина кратчайшего пути в диапазоне от 4 до 6 и диаметр порядка 10, т.е. в тех графах, которые по этим параметрам наиболее близки к реальным социальным графам.

Algorithm 1: алгоритм Флажолле-Мартена для расчета центральности вершин графа

```

Input: граф из N вершин  $V\{v_1, v_2, \dots, v_n\}$ 
матрица сопряженности  $g\{0,1\}^{N \times N}$ 
коэффициент распада  $\delta \leftarrow const(0 : 1)$ 
Output: расчетные значения центральности
распада  $C^{dec}\{c_1^{dec}, c_2^{dec}, \dots, c_n^{dec}\}$ 
for v in V do
    //обнуление центральности вершины v
     $c_v^{dec} \leftarrow 0;$ 
     $i \leftarrow 0;$ 
    // инициализация кодовой строки
    while  $i < \log_2 N$  do
         $M_{prev}(v)[i] \leftarrow 0;$ 
         $M_{prev}(v)[i] \leftarrow 1$  с вероятностью  $p=1/2^{i+1}$ 
         $i \leftarrow i + 1$ 
    end
     $b \leftarrow$  номер первого нулевого разряда в  $M_{prev}(v);$ 
    // условное число соседей вершины на 0-й
    итерации
     $N_v(r) \leftarrow 2^b / 0,77359;$ 
end
//флаг достижения максимального числа итераций
flag  $\leftarrow 0;$ 
r  $\leftarrow 0;$ 
Итерация алгоритма:
while flag = 0 do
    // счетчик расстояния от вершины v
    r  $\leftarrow r + 1;$ 
    for v in V do
        // логическое ИЛИ кодовых строк вершины
        v и ее соседей  $\{u|g(v,u)=1\}$ 
         $M_{cur}(v) \leftarrow M_{prev}(v) \oplus \{M_{prev}(u)|g(v,u)=1\};$ 
        // проверка на достижение максимального
        числа итераций:
        if  $M_{cur}(v) = M_{prev}(v)$  then
            | flag  $\leftarrow 1;$ 
        else
            | flag  $\leftarrow 0;$ 
        end
         $b \leftarrow$  номер первого нулевого разряда в
         $M_{cur}(v)$ 
        //расчетное число соседей вершины на
        расстоянии r
         $N_v(r) \leftarrow 2^b / 0,77359 - N_v(r);$ 
        //значение центральности вершины v на
        текущей итерации
         $c_v^{dec} \leftarrow c_v^{dec} + N_v(r) * \delta^r;$ 
    end
    for v in V do
        |  $M_{prev}(v) \leftarrow M_{cur}(v)$ 
    end
end
end

```

Рис.2. Адаптированный алгоритм Флажолле-Мартена для расчета центральности распада вершин графа.

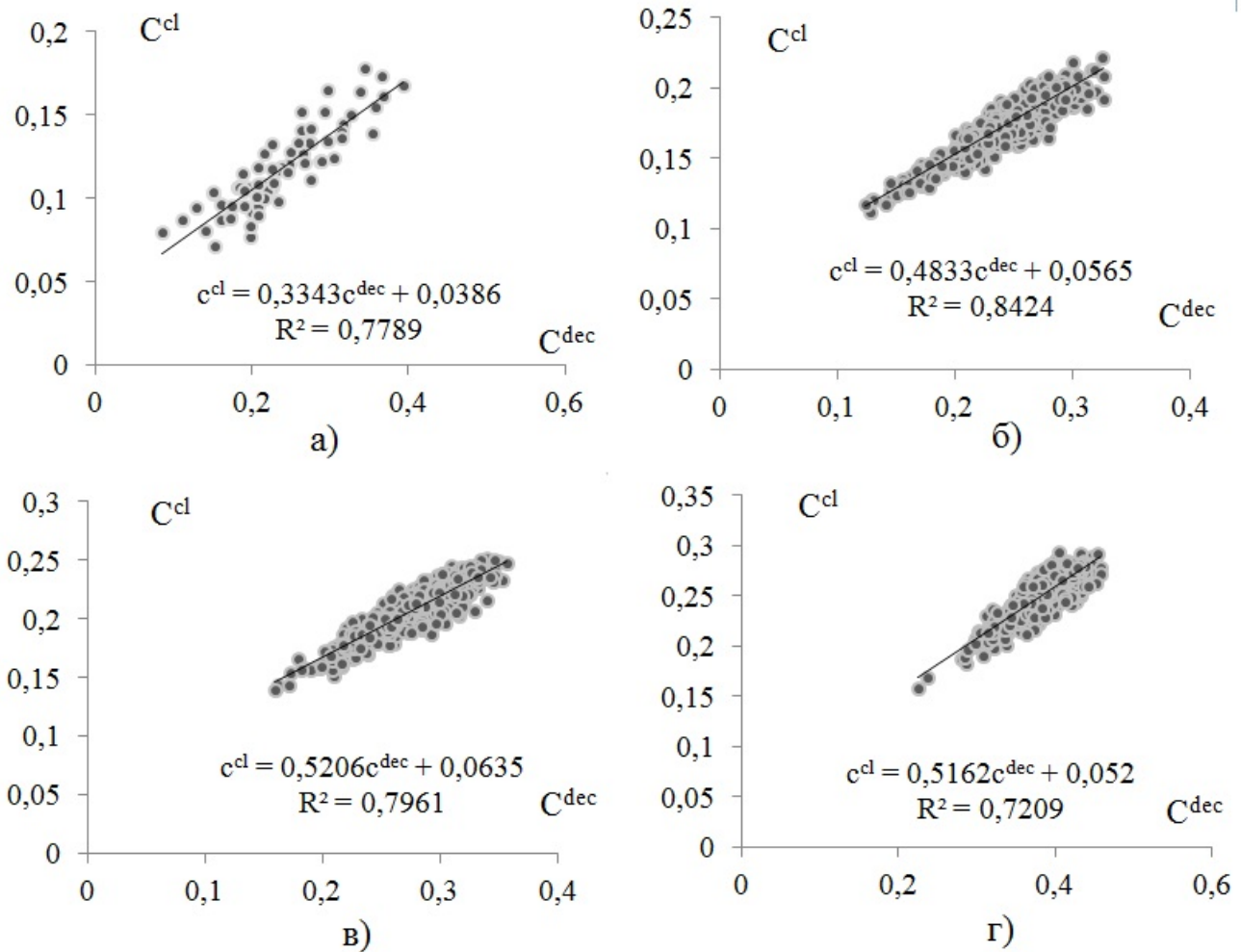


Рис.3. Диаграммы разброса центральности по близости  $C_V^{cl}$  (по оси y) и центральности распада  $C_V^{dec}$  (по оси x) с коэффициентом распада  $\delta = 0,8$  для  $G(n,p)$ -графов с параметрами:

- а)  $V = 500, p = 0,002, V_c = 73, k_{clust} = 0, \bar{l} = 8.89, d = 22;$
- б)  $V = 500, p = 0,005, V_c = 464, k_{clust} = 0,002, \bar{l} = 5.96, d = 13;$
- в)  $V = 00, p = 0,0075, V_c = 485, k_{clust} = 0,003, \bar{l} = 4.87, d = 10;$
- г)  $V = 500, p = 0,01, V_c = 498, k_{clust} = 0.006, \bar{l} = 4.09, d = 9$

В завершение эксперимента аналогичным образом был проанализирован социальный граф, основанный на реальных данных, полученных из онлайн-социальной сети Facebook. Эти данные<sup>1</sup> получены авторами работы [17]. В их основе лежат 10 пользовательских графов Facebook (ego network), каждый из которых состоит из друзей одного конкретного пользователя и связей между ними. Затем все 10 графов объединены на основе содержащихся в них общих друзей и связей между ними (рис. 4).

Число вершин итогового графа  $V = 4039$ , граф связный, т.е. между каждой парой вершин имеется путь, число ребер графа насчитывает  $E = 88234$ . Этот реальный социальный граф отличается от рассмотренных выше случайных  $G(n,p)$ -графов высоким коэффициентом кластеризации, которая в предыдущих

примерах практически отсутствовала. Диаграмма разброса  $C_V^{dec}$  и  $C_V^{cl}$  на рисунке 5.

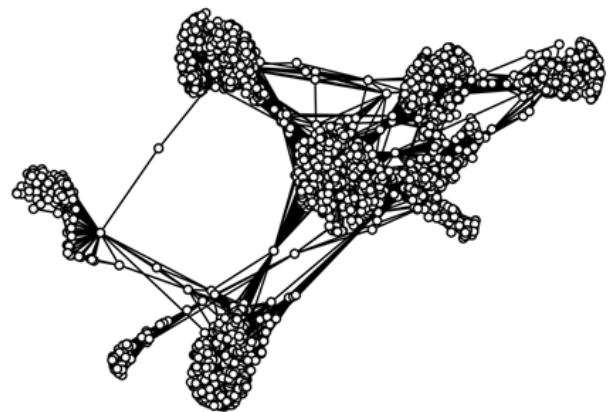


Рис.4. Объединенный граф, состоящий из 10 пользовательских графов онлайн-социальной сети Facebook.

<sup>1</sup> <https://snap.stanford.edu/data/egonets-Facebook.html>



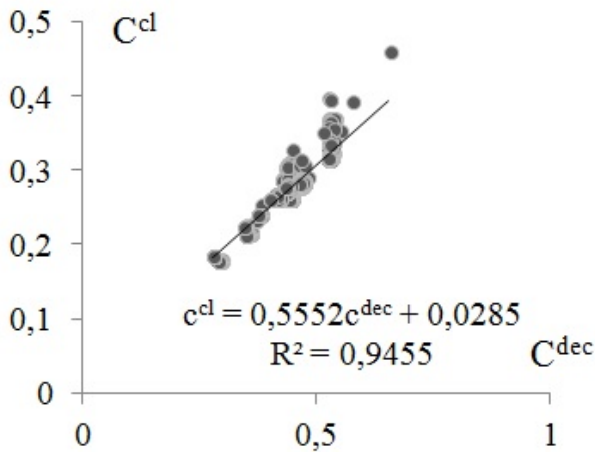


Рис.5. Диаграмма разброса центральности по близости  $c_v^{cl}$  (по оси y) и центральности распада  $c_v^{dec}$  (по оси x) с коэффициентом распада  $\delta = 0,8$  для графа Facebook с параметрами:

$$V = 4039, p = 0,015, V_c = 4039, k_{clust} = 0.606, \bar{l} = 3.69, d = 8$$

## VI. ВЫВОДЫ

Результаты расчета, в том числе, на реальных данных сети Facebook, показывают высокую взаимозависимость между  $c_v^{dec}$  и  $c_v^{cl}$ , что позволяет сделать заключение: две этих метрики центральности отражают одну и ту же характеристику вершин социального графа, связанную с длинами кратчайших путей. При этом центральность распада – это более гибкий инструмент, нежели центральность по близости. Она позволяет при анализе графа, манипулируя значением коэффициента распада  $\delta$ , регулировать значимость вершин, находящихся на разном удалении от вершины  $v$ , тем самым объединяя центральность по степени и центральность по близости в разных пропорциях.

Предложенный метод расчета центральности распада в обобщенной форме на основе алгоритма Флажолемартена позволяет добиться существенного выигрыша в скорости вычислений. По сравнению с алгоритмом Дейкстры преимущество можно оценить как  $(VE + V^2 \times \log_2 V) / dE$ , что в наихудшем случае, при  $E \rightarrow V^2$ , составит для больших графов с высокой плотностью связей порядка  $V$ .

## БИБЛИОГРАФИЯ

- [1] Freeman L.C. Centrality in social networks: Conceptual clarification // *Social Networks*. 1978. №1. P. 215-239.
- [2] Tsvetovat M. Kouznetsov A. *Social Network Analysis for Startups*. O'Reilly Media Inc., 2011. 192 p.
- [3] Banerjee A., Chandrasekhar A.G., Duflo E., Jackson M.O. The Diffusion of Microfinance // *Science*. 2013. №. 341 P. 363-340. DOI:10.1126/science.1236498.
- [4] Kang U., Papadimitriou S., Sun J., Tong H. Centralities in Large Networks: Algorithms and Observations // *Proceedings of the 2011 SIAM International Conference on Data Mining*. 2011. P. 119-130.

- [5] Johnson D. B. Efficient algorithms for shortest paths in sparse networks // *Journal of the ACM*. 1977. V.24. №1. P. 1-13.
- [6] Zhao J., John C.S. Lui, Towsley D., Guan X.H. Measuring and Maximizing Group Closeness Centrality over Disk-Resident Graphs // *Proceedings of the 23rd International Conference on World Wide Web. Republic and Canton of Geneva, Switzerland*. 2014. P. 689–694.
- [7] Jackson M.O., Wolinsky A. A Strategic Model of Social and Economic Networks // *Journal of Economic Theory*. 1996. V. 71. №1 P. 44 – 74.
- [8] Jackson M.O. *Social and Economic Networks*. Princeton University Press, 2008. 520 p.
- [9] Milgram S. The Small-World Problem // *Psychology Today*. 1967. №1. P. 60–67.
- [10] Adamic L.A. The SmallWorldWeb // *Proceedings of the International Conference on Theory and Practice of Digital Libraries*. Springer Berlin Heidelberg. 1999. P. 443-452.
- [11] Cannarella J., Spechler J.A. Epidemiological modeling of online social network dynamics. 2014. Cornell University Library. <https://arxiv.org/abs/1401.4208>
- [12] Backstrom, L., P. Boldiy, M. Rosay, J. Ugander S. Vignay. Four Degrees of Separation. 2012. Cornell University Library. <https://arxiv.org/abs/1111.4570>
- [13] Catanese S., De Meo P., Ferrara E., Fiumara G. Analyzing the Facebook friendship graph // *CEUR Workshop Proceedings*. 2010. P.14-19
- [14] Flajolet P., Martin G.N. Probabilistic counting algorithms for data base applications // *Journal of Computer and System Sciences*. 1985. V.31. №2. P. 182–209
- [15] Palmer C.R., Gibbons P.B., Faloutsos C. ANF: A Fast and Scalable Tool for Data Mining in Massive Graphs // *Proceedings of the Eighth ACM SIGKDD international Conference on Knowledge Discovery and Data Mining*. 2002. P. 81-90.
- [16] Erdos P., Renyi A. On Random Graphs. I. // *Publicationes Mathematicae*. 1959. V.6. P.290–297.
- [17] McAuley J., Leskovec J. Learning to Discover Social Circles in Ego Networks // *Proceedings of the Advances in Neural Information Processing Systems 25*. 2012. PP. 548–556.

## **Decay centrality in social graphs and Flajolet-Martin algorithm adaptation for its computation**

B. Toropov

*Abstract*– The paper is devoted to a node centrality in a graph, which is a generalization of degree and closeness centralities for social graphs. This centrality form is based on a power function, known as decay centrality and proposed by Jackson and Wolinsky. Author examines Flajolet-Martin adopted algorithm for decay centrality approximation. The results are compared to the real closeness centrality values. Computational experiment has shown, that centrality form exposed along with the algorithm adopted for its computation are suited for social network graphs closeness centrality approximation. Research methodology is based on elements of graph theory and social network analysis.

*Keywords*– social graph, social network, centrality, Flajolet-Martin algorithm, approximation.

Toropov Boris  
assistant professor of Management Academy of the Ministry of the Interior of Russia (e-mail: torbor@mail.ru)