

Оптимизация, системный анализ и исследование операций

© 2025 г. М.А. ТУЖИЛИН, канд. физ.-мат. наук (mtu93@mail.ru)

(Национальный исследовательский университет

«Высшая школа экономики», Москва;

Московский государственный университет им. М.В. Ломоносова)

СООТНОШЕНИЯ МЕЖДУ ДЛИНОЙ СРЕДНЕГО КРАТЧАЙШЕГО ПУТИ И ДРУГИМИ ЦЕНТРАЛЬНОСТЯМИ В ГРАФАХ

Доказываются теоремы о связи между длиной среднего кратчайшего пути в простых связных графах и радиальной центральностью, центральностью по близости, центральностью напряжения и средним кластерным коэффициентом.

Ключевые слова: сети, центральности, локальные и глобальные характеристики графов, средний кратчайший путь, средний кластерный коэффициент.

DOI: 10.31857/S0005231025050072, **EDN:** AWYRUY

1. Введение

Длина среднего кратчайшего пути является одной из самых важных характеристик графов реального мира (или сетей). Первые попытки вычисления этой характеристики были предприняты в 1969 г. американскими социальными психологами Стэнли Милгрэмом и Джефффри Трэверсом для графа знакомых [1]. Они обнаружили, что каждый человек опосредованно знаком с любым другим человеком через цепочку из шести знакомств. Эта теория в последствии была названа теорией шести рукопожатий. Далее эта теория была доказана Дунканом Уоттсом и Стивеном Строгацем в статье [2]. Оказалось, что многие графы, получающиеся в реальных прикладных задачах, имеют похожее свойство: имеют небольшую длину среднего кратчайшего расстояния.

Другой характеристикой графов в прикладных задачах является наличие большого среднего кластерного коэффициента или кластерного коэффициента Уоттса и Строгаца [2]. Графы с такими характеристиками (небольшой длиной среднего кратчайшего расстояния и большим средним кластерным коэффициентом) называются сетями малого мира. Кроме длины среднего кратчайшего расстояния и среднего кластерного коэффициента, в теории социальных сетей также выделяют ряд других важных характеристик, которые называются центральностями (например, центральность напряжения, близости, посредничества, радиальная центральность и пр.). Для того чтобы понять, какими внутренними характеристиками обладает граф, полученный в приложениях, анализируются распределения этих центральностей

(см., например, [3–7]), однако наличие зависимостей между центральностями является мало изученным вопросом. В данной статье доказываются теоремы о зависимости между длиной среднего кратчайшего пути и радиальной центральностью, центральностью по близости, центральностью напряжения, а также было получено соотношение между длиной среднего кратчайшего пути и средним кластерным коэффициентом для геодезических графов.

2. Основные определения

Все последующие определения даются для простого связного неориентированного графа G . Введем необходимые обозначения. Обозначим:

- $V(G)$ – множество вершин графа, $E(G)$ – множество ребер графа, $A = \{a_{ij}\}$ – матрица смежности графа G ,
- $N(v)$ – индуцированный подграф в графе G на вершинах, смежных с вершиной v ,
- $N'(v)$ – индуцированный подграф в графе G на вершинах $V(N(v)) \cup \{v\}$,
- $\bar{f}(x_1, \dots, x_k)$ для любой функции $f : V \times V \times \dots \times V \rightarrow \mathbb{R}$ – ограничение этой функции на подграф $N'(v)$ (например $\bar{L}(x, y)$ – среднее кратчайшее расстояние между вершинами x и y в подграфе $N'(v)$),
- $d_i = \deg(v_i)$,
- $\text{dist}(s, t)$ – длина кратчайшего расстояния между s и t ,
- $n = |V(G)|$, $m = |E(G)|$,
- $X(i) = X(v_i)$ для любого X – функции или множества, соответствующего вершине v_i .

Дадим определения центральностей.

1. Диаметр графа $\text{diam}(G) = \max_{s, t \in V(G)} \text{dist}(s, t)$.

2. Длина среднего кратчайшего пути в графе

$$L(G) = \frac{1}{n(n-1)} \sum_{s, t \in V(G), s \neq t} \text{dist}(s, t).$$

3. Локальный кластерный коэффициент

$$c_i = c(i) = \frac{\text{число ребер в подграфе } N(i)}{\text{максимально возможное число ребер в подграфе } N(i)} = \frac{2|E(N(i))|}{d_i(d_i-1)}.$$

4. Средний кластерный коэффициент графа

$$C_{WS}(G) = \frac{1}{n} \sum_{i \in V(G)} c_i = \frac{1}{n} \sum_{i \in V(G)} \frac{2|E(N(i))|}{d_i(d_i-1)} = \frac{1}{n} \sum_{i \in V(G)} \frac{\sum_{j, k \in V(G)} a_{ij} a_{jk} a_{ki}}{d_i(d_i-1)}.$$

5. Центральность по близости $\text{Clo}(v) = \frac{n-1}{\sum_{t \in V(G)} \text{dist}(v, t)}$.

6. Радиальная центральность $\text{Rad}(v) = \frac{\sum_{t \in V(G), t \neq v} (\text{diam}(G)+1-\text{dist}(v, t))}{n-1}$.

7. Центральность напряжения $\text{Str}(i) = \sum_{s, t \in V(G), s \neq t \neq i} \sigma_{st}(i)$, где $\sigma_{st}(i)$ – число кратчайших путей из s в t через вершину i .

Заметим, что все центральности неотрицательные, а также некоторые из них меньше либо равны 1, а именно: $c_i, C_{WS}, \text{Clo}(v)$.

Далее потребуется

Определение 1. Граф называется геодезическим, если между любыми двумя вершинами существует единственный кратчайший путь.

Это определение эквивалентно условию, когда в графе могут быть только циклы нечетной длины.

3. Основные результаты

В общем случае индуцированный подграф $G' \subset G$ не обязан быть связным. В таком случае можно определить длину среднего кратчайшего расстояния между вершинами подграфа G' по отношению к расстоянию dist в объемлющем графе G . Назовем $L(N(i))$ локальной длиной среднего кратчайшего расстояния для вершины i .

Начнем с доказательства некоторых несложных соотношений: докажем лемму о связи между локальной длиной среднего кратчайшего расстояния и локальным кластерным коэффициентом этой вершины.

Лемма 1.

$$L(N(i)) = 2 - c_i.$$

Доказательство.

$$\begin{aligned} L(N(i)) &= \frac{1}{d_i(d_i - 1)} \sum_{s, t \in N(i), s \neq t} \text{dist}(s, t) = \\ &= \frac{1}{d_i(d_i - 1)} \sum_{(s, t) \in E(N(i))} \text{dist}(s, t) + \sum_{s, t \in N(i), (s, t) \notin E(N(i))} \text{dist}(s, t) = \\ &= \frac{1}{d_i(d_i - 1)} \left(2|E(N(i))| + \sum_{(s, i), (i, t) \in E(G), (s, t) \notin E(G)} \text{dist}(s, t) \right) = \\ &= \frac{1}{d_i(d_i - 1)} \left(2|E(N(i))| + 2(d_i(d_i - 1) - 2|E(N(i))|) \right) = 2 - c_i. \end{aligned}$$

Отметим, что кратчайшее расстояние для вершин из $N(i)$ считается по всему графу G .

Усреднением по всем вершинами получаем тривиальное следствие о связи между средним кластерным коэффициентом и средней локальной длиной среднего кратчайшего расстояния.

Следствие 1.

$$C_{WS}(G) = 2 - \frac{1}{n} \sum_{i \in V(G)} L(N(i)).$$

Докажем лемму о связи длины среднего кратчайшего расстояния в подграфе $N'(i)$ с длиной среднего кратчайшего расстояния в $N(i)$.

Лемма 2.

$$L(N'(i)) = \frac{(d_i - 1)L(N(i)) + 2}{d_i + 1}.$$

Доказательство. Распишем определение

$$L(N'(i)) = \frac{1}{(d_i + 1)d_i} \sum_{s, t \in V(N'(i)), s \neq t} \text{dist}(s, t) = \frac{d_i - 1}{d_i + 1} L(N(i)) + \frac{2}{d_i + 1}.$$

Докажем теорему о связи длины среднего кратчайшего расстояния в индуцированном подграфе с длиной среднего кратчайшего расстояния в объемлющем графе, если индуцированный подграф получен из объемлющего графа удалением одной вершины.

Теорема 1. Пусть граф G получен из простого связного графа G' удалением одной вершины и $|V(G)| = n$. Тогда

$$L(G') \geq \frac{n}{n + 1} L(G),$$

где длина среднего кратчайшего расстояния $L(G)$ определена в объемлющем графе G' , если G не связный.

Доказательство. Обозначим удаленную вершину через v . По неравенству треугольника имеем $\forall s, t \in V(G) : \text{dist}(s, v) + \text{dist}(v, t) \geq \text{dist}(s, t)$, где равенство достигается, когда не существует пути из s в t в графе G . Следовательно,

$$\begin{aligned} \sum_{s, t \in V(G), s \neq t} (\text{dist}(s, v) + \text{dist}(v, t)) &\geq \sum_{s, t \in V(G), s \neq t} \text{dist}(s, t), \\ \frac{2(n - 1)}{n(n - 1)} \sum_{t \in V(G)} \text{dist}(v, t) &\geq \frac{1}{n(n - 1)} \sum_{s, t \in V(G), s \neq t} \text{dist}(s, t), \\ \frac{2}{n} \sum_{t \in V(G)} \text{dist}(v, t) &\geq L(G). \end{aligned}$$

Тогда,

$$\begin{aligned} L(G') &= \frac{1}{(n + 1)n} \sum_{s, t \in V(G'), s \neq t} \text{dist}(s, t) = \\ &= \frac{1}{(n + 1)n} \left(2 \sum_{t \in V(G)} \text{dist}(v, t) + \sum_{s, t \in V(G), s \neq t} \text{dist}(s, t) \right) = \\ &= \frac{1}{n + 1} \frac{2}{n} \sum_{t \in V(G)} \text{dist}(v, t) + \frac{n - 1}{n + 1} L(G) \geq \frac{n}{n + 1} L(G). \end{aligned}$$

Заметим, что если G состоит из n изолированных вершин, то неравенство становится равенством.

Получаем следствия.

Следствие 2. Пусть граф G получен из графа $G' \subset H$ удалением одной вершины и $|V(G)| = n$, граф H связный и простой. Тогда

$$L(G') \geq \frac{n}{n+1} L(G),$$

где длины средних кратчайших расстояний $L(G)$ и $L(G')$ определены в объемлющем графе H , если соответствующие графы не связные.

Доказательство. Доказательство аналогичное, как в предыдущей теореме, и равенство также достигается в случае, когда G состоит из изолированных вершин.

Следствие 3.

$$L(N'(i)) \geq \frac{d_i}{d_i+1} L(N(i)).$$

Теперь докажем теорему о связи длины среднего кратчайшего расстояния в индуцированном подграфе с длиной среднего кратчайшего расстояния в объемлющем графе.

Теорема 2. Рассмотрим индуцированный подграф $G' \subset G$. Пусть $|V(G)| = n$, $|V(G')| = n + k$. Тогда

$$L(G') \geq \frac{n}{n+k} L(G),$$

где длина среднего кратчайшего расстояния $L(G)$ определена в объемлющем графе G' , если G не связный.

Доказательство. Рассмотрим процесс получения графа G' из графа G последовательным добавлением k вершин и соответствующих ребер. Сначала будем последовательно добавлять вершины, смежные с вершинами графа G . Добавляя по одной вершине, получим последовательность графов $G_1, G_2, \dots, G_p$, где $|V(G_i)| = n + i$. Далее будем последовательно добавлять вершины, смежные с вершинами графа G_p , и получим последовательность $G_{p+1}, G_{p+2}, \dots$. В конечном итоге получим граф G' . По предыдущему следствию имеем

$$\begin{aligned} L(G') &\geq \frac{n+k-1}{n+k} L(G_{k-1}) \geq \frac{n+k-1}{n+k} \frac{n-k-2}{n-k-1} L(G_{k-2}) = \\ &= \frac{n-k-2}{n+k} L(G_{k-2}) \geq \dots \geq \frac{n}{n+k} L(G). \end{aligned}$$

Докажем лемму о связи средней центральности по близости и среднего кратчайшего расстояния в графе.

Лемма 3.

$$L(G) \geq \frac{n}{\sum_{v \in V(G)} \text{Clo}(v)}.$$

Доказательство. Воспользуемся неравенством о среднем гармоническом и арифметическом:

$$\frac{1}{n} \sum_{v \in V(G)} \text{Clo}(v) = \frac{1}{n} \sum_{v \in V(G)} \frac{n-1}{\sum_{t \in V(G)} \text{dist}(v, t)} \geq \frac{n(n-1)}{\sum_{v, t \in V(G)} \text{dist}(v, t)} = \frac{1}{L(G)}.$$

Заметим, что равенство выполнено, когда все средние кратчайшие расстояния от каждой вершины до всех других равны.

Докажем лемму о связи длины среднего кратчайшего расстояния в графе со средней радиальной центральностью.

Лемма 4.

$$L(G) = \text{diam}(G) + 1 - \frac{1}{n} \sum_{v \in V(G)} \text{Rad}(v).$$

Доказательство. Доказательство следует напрямую из определения:

$$\begin{aligned} \frac{1}{n} \sum_{v \in V(G)} \text{Rad}(v) &= \frac{1}{n} \sum_{v \in V(G)} \frac{(n-1)(\text{diam}(G) + 1) - \sum_{t \in V(G), t \neq v} \text{dist}(v, t)}{n-1} = \\ &= \text{diam}(G) + 1 - L(G). \end{aligned}$$

Докажем теорему о связи между длиной среднего кратчайшего расстояния в графе и средней центральностью напряжения для геодезических графов.

Теорема 3. Если G – геодезический граф, то

$$L(G) = 1 + \frac{1}{n(n-1)} \sum_{i \in V(G)} \text{Str}(i).$$

Доказательство. Определим функцию

$$\chi_{st}(i) = \begin{cases} 1, & \text{если } i \neq s \neq t - \text{вершина кратчайшего пути между } s \text{ и } t, \\ 0, & \text{в остальном.} \end{cases}$$

Следовательно, для центральности напряжения $\text{Str}(i) = \sum_{s, t \in V(G)} \chi_{st}(i)$.

Если в графе G между любыми двумя вершинами существует единственный кратчайший путь, то $\text{dist}(s, t) = \sum_{i \in V(G)} \chi_{st}(i) + 1$ (в противном случае

$\text{dist}(s, t) \leq \sum_{i \in V(G)} \chi_{st}(i) + 1$. Следовательно, для любого i

$$\begin{aligned} \sum_{s, t \in V(G)} \text{dist}(s, t) &= 2|E| + \sum_{s, t \in V(G), \text{dist}(s, t) \geq 2} \text{dist}(s, t) = \\ &= 2|E| + \sum_{s, t \in V(G), \text{dist}(s, t) \geq 2} \left(\sum_{i \in V(G)} \chi_{st}(i) + 1 \right) = \\ &= 2|E| + \sum_{i \in V(G)} \text{Str}(i) + n(n-1) - 2|E| = \sum_{i \in V(G)} \text{Str}(i) + n(n-1). \end{aligned}$$

Получаем

Следствие 4. Для любого просто связного графа G

$$L(G) \leq 1 + \frac{1}{n(n-1)} \sum_{i \in V(G)} \text{Str}(i).$$

Для получения соотношения между длиной среднего кратчайшего расстояния и средним кластерным коэффициентом сначала докажем теорему о связи среднего кластерного коэффициента с центральностью напряжения.

Теорема 4. Для простого связного графа G без висячих вершин

$$C_{WS}(G) \geq \frac{1}{n} \sum_{i \in V(G)} \left(1 - \frac{\text{Str}(i)}{d_i(d_i-1)} \right).$$

Доказательство. Заметим, что $\forall j, k \in N(i) : (j, k) \notin E(N(i))$ кратчайшее расстояние между j и k – это $j \rightarrow i \rightarrow k$. Тогда

$$\begin{aligned} \text{Str}(i) &\geq 2 \left(\frac{d_i(d_i-1)}{2} - |E(N(i))| \right), \\ \frac{1}{d_i(d_i-1)} \text{Str}(i) &\geq 1 - c_i. \end{aligned}$$

Усреднением по i получаем:

$$C_{WS}(G) \geq \frac{1}{n} \sum_{i \in V(G)} \left(1 - \frac{\text{Str}(i)}{d_i(d_i-1)} \right).$$

Заметим, что равенство достигается, если $\text{diam}(G) = 2$.

Следствие 5. Для любого простого связного графа G

$$C_{WS}(G) \geq \frac{1}{n} \sum_{i \in V(G)} \left(1 - \frac{\text{Str}(i)}{d_i(d_i-1)} \right) - \frac{\text{число висячих вершин}}{n}.$$

Доказательство. Для каждой висячей вершины положим $\frac{\text{Str}(i)}{d_i(d_i-1)} = 0$, тогда в правой части неравенства $\frac{1}{d_i(d_i-1)}\text{Str}(i) \geq 1 - c_i$ будет 1, а в левой 0. Если добавим в левую часть 1 для каждой висячей вершины, то получим верное равенство.

Теперь докажем теорему о связи между длиной среднего кратчайшего расстояния и средним кластерным коэффициентом.

Теорема 5. Если G – геодезический граф и $\forall i, j \in V(G)$ выполнено, если $d_i \leq d_j$, то $\text{Str}(i) \leq \text{Str}(j)$, тогда

$$1 - C_{WS}(G) \leq \frac{1}{n} \sum_{i \in V(G)} \frac{(L(G) - 1)(n - 1)}{d_i(d_i - 1)} + \frac{\text{число висячих вершин}}{n}.$$

Доказательство. Перенумеруем вершины так, что $\forall i \leq j : d_i \leq d_j$. Тогда для $i \leq j$ выполнено $\text{Str}(i) \leq \text{Str}(j)$ и $d_i(d_i - 1) \leq d_j(d_j - 1)$. По теореме 3 и следствию 5, если в графе нет висячих вершин, получаем

$$\begin{aligned} 1 - C_{WS}(G) &\leq \frac{1}{n} \sum_{i \in V(G)} \frac{\text{Str}(i)}{d_i(d_i - 1)} \stackrel{\text{неравенство Чебышева}}{\leq} \\ &\stackrel{\text{неравенство Чебышева}}{\leq} \left(\frac{1}{n} \sum_{i \in V(G)} \text{Str}(i) \right) \left(\frac{1}{n} \sum_{i \in V(G)} \frac{1}{d_i(d_i - 1)} \right) = \\ &= \frac{1}{n} \sum_{i \in V(G)} \frac{(L(G) - 1)(n - 1)}{d_i(d_i - 1)}. \end{aligned}$$

В случае висячих вершин справа добавится слагаемое $\frac{\text{число висячих вершин}}{n}$. Отметим, что если существуют две вершины $i, j \in V(G)$ такие, что $d_i < d_j$ и $\text{Str}(i) < \text{Str}(j)$, то неравенство в этой теореме будет строгое.

Пример 1. Рассмотрим граф звезды с $|V(G)| = n + 1$ вершиной. Это геодезический граф. Центральная вершина имеет степень n , локальный кластерный коэффициент $c_i = 0$ и центральность стресса $\text{Str}(i) = n(n - 1)$. Все остальные вершины висячие ($d_i = 1$, $c_i = 0$, $\text{Str}(i) = 0$), поэтому для этого графа выполнено условие теоремы 5.

$$\begin{aligned} L(G) &= \frac{n(2n - 1) + n}{(n + 1)n} = \frac{2n}{n + 1}, \quad C_{WS} = 0. \\ 1 - C_{WS}(G) &= 1 = \frac{\frac{n-1}{n+1}n}{n(n-1)} + \frac{n}{n+1} = \\ &= \frac{1}{n+1} \sum_{i \in V(G)} \frac{(L(G) - 1)n}{d_i(d_i - 1)} + \frac{\text{число висячих вершин}}{n+1}. \end{aligned}$$

Следовательно, для этого графа выполнено равенство.

СПИСОК ЛИТЕРАТУРЫ

1. *Travers J., Milgram S.* An experimental study of the small world problem / Social networks. Academic Press. 1977. P. 179–197.
2. *Watts D.J., Strogatz S.H.* Collective dynamics of ‘small-world’ networks // Nature. 1998. V. 393. No. 6684. P. 440–442.
3. *Borgatti S.P., Everett M.G.* A graph-theoretic perspective on centrality // Social networks. 2006. V. 28. No. 4. P. 466–484.
4. *Kiss C., Bichler M.* Identification of influencers-measuring influence in customer networks // Decision Support Systems. 2008. V. 46. No. 1. P. 233–253.
5. *Lee S.H.M., Cotte J., Noseworthy T.J.* The role of network centrality in the flow of consumer influence // J. Consum. Psychol. 2010. V. 20. No. 1. P. 66–77.
6. *You J., Gomes-Selman J.M., Ying R., et al.* Identity-aware graph neural networks // Proceedings of the AAAI conference on artificial intelligence. 2021. V. 35. No. 12. P. 10737–10745.
7. *Yuan M.M., Guo X., Wu L., et al.* Climate warming enhances microbial network complexity and stability // Nature Climate Change. 2021. V. 11. No. 4. P. 343–348.

Статья представлена к публикации членом редколлегии А.Г. Кушнером.

Поступила в редакцию 16.12.2024

После доработки 30.12.2024

Принята к публикации 09.01.2025