Математические структуры и моделирование
2010, вып. 21, с. 5–10

УДК 519.17

МЕТРИЧЕСКАЯ СВЯЗНОСТЬ ВЕРШИН ГРАФА

С.В. Белим, В.Ю. Бардычев

В данной работе проводится исследование одного из возможных вариантов функции, определяющей меру связности вершин взвешенного графа. Исследования проводятся на основе компьютерного эксперимента.

Введение

При исследовании графов важную роль играет понятие связности, понимаемое как существование пути из одной вершины графа в другую. Однако в ряде прикладных задач важным становится не только сама достижимость из одной вершины другой вершины, но и длина соответствующего пути.

Будем считать, что дуги определяют некоторые взаимодействия или связи между вершинами. Поставим себе задачу нахождения не только непосредственных связей между вершинами, но и связи через посредников, то есть более длинные пути. В качестве приложений данной проблемы можно привести квантовые системы, в которых роль множества вершин играют состояния системы, а множества дуг – переходы между состояниями. При этом возможны переходы как напрямую между состояниями, так и через цепочку промежуточных состояний. Другим примером является связь между научными работами по одной тематике. Подавляющее количество научных работ основывается на результатах, полученных в более ранних работах. Причем авторы отдельно взятой работы могут основываться на результатах других авторов непосредственно, а могут через цепочку промежуточных результатов третьих авторов.

1. Постановка задачи

Пусть граф определен множеством вершин V и множеством дуг D. Отношение инцидентности будем задавать с помощью матрицы смешения E (mixing matrix) [1]. А именно, если существует дуга из вершины v_i в вершину v_j, то соответствующий элемент матрицы смешения $E_{ij} = 1$, в противном случае $E_{ij} = 0$. Также будем считать, что в графе нет петель, и полагать $E_{ii} = 0$.

Попытаемся получить числовую характеристику, показывающую степень связности двух вершин. Причем это не просто количество вершин или дуг, удаление которых к тому, что заданные вершины попадают в разные компоненты

Copyright © 2010 **С.В. Белим, В.Ю. Бардычев.**
Омский государственный университет им. Ф.М. Достоевского.
E-mail: sbelim@mail.ru

связности. Наша оценка должна учитывать длины путей, соединяющих эти вершины. Соответствующую числовую характеристику будем называть метрической связностью.

Пусть между вершинами v_i и v_j некоторого графа существует несколько путей длины k. Обозначим количество таких путей через $p_k(v_i, v_j)$. В общем случае будем считать, что пути разной длины дают разный вклад в метрическую связность двух вершин. Введем функцию $g(k)$, которую в дальнейшем будем называть весовой функцией пути. Более точно $g(k)$ показывает вклад в связь вершин пути длиной k. Тогда метрическая связность двух вершин может быть определена следующим образом:

$$w(v_i, v_j) = \sum_{k=1}^{\infty} g(k) p_k(v_i, v_j).$$

Рассматривая граф как представление влияния элементов множества V друг на друга можно сказать, что наличие дуги между вершинами показывает прямое влияние, путь длины 2 – влияние через одного посредника, длины 3 – через двух посредников и так далее, длины n — через $n-1$ посредника.

Определим матрицу метрической связности вершин M, показывающую взаимное влияние всех вершин графа друг на друга:

$$M_{ij} = w(v_i, v_j).$$

Как показано в работе [2], матрица метрической связности может быть вычислена на основе матрицы смешения:

$$M = \sum_{k=1}^{\infty} g(k) E^k.$$

Матрица метрической связности играет роль, аналогичную матрице достижимости [2], служащей для определения связности графа.

При определении метрической связности важным является вопрос выбора весовой функции пути $g(k)$. Рассмотрим свойства, которым должна удовлетворять эта функция:

1. $g(1) = 1$, то есть наличие дуги между двумя вершинами должно давать единичный вклад в метрическую связность.

2. $g(k) > g(k+1)$, для всех $k \in [1, \infty)$. То есть функция $g(k)$ должна быть убывающей.

3. $g(k) > 0$, для всех $k \in [1, \infty)$. Наличие путей должно увеличивать метрическую связность.

Рассмотрим два возможных варианта выбора функции $g(k)$.

1. $g(k) = 1/\alpha^{k-1}$, где $0 < \alpha < 1$. Функция данного вида была рассмотрена в [1] для выявления связных структур методом построения иерархического дерева. В этом случае метрическая связность двух вершин вычисляется по формуле

$$w(v_i, v_j) = \sum_{k=1}^{\infty} \alpha^{k-1} p_k(v_i, v_j).$$

Матрица метрической связности имеет вид:

$$M = \sum_{k=1}^{\infty} \alpha^{k-1} E^k = \alpha^{-1}((I - \alpha E)^{-1} - I).$$

Здесь I – единичная матрица. Коэффициент α имеет двоякий смысл. С одной стороны, он показывает, во сколько раз уменьшается относительное влияние пути при увеличении его длины на единицу. С другой стороны, рассмотрим ситуацию, при которой из вершины v_i в вершину v_j ведет несколько путей одинаковой длины. В частности, если имеется $p_2(v_i, v_j) = 1/\alpha$ путей длины 2, то они дают такой же вклад в метрическую связность, как и дуга между вершинами. При наличии $p_3(v_i, v_j) = 1/\alpha$ путей длины 3 их вклад будет эквивалентен одному пути длиной 2. Если же имеется $p_3(v_i, v_j) = 1/^2$ путей длины 3, то их вклад эквивалентен одной дуге. В общем случае при наличии $p_n(v_i, v_j) = 1/\alpha^{n-k}$ путей длины n их вклад будет эквивалентен одному пути длины k. Эти соображения позволяют выбирать коэффициент α из практических соображений при постановке задачи.

2. $g(k) = 1/k$. Метрическая связность двух вершин будет вычисляться по формуле

$$w(v_i, v_j) = \sum_{k=1}^{\infty} k^{-1} p_k(v_i, v_j).$$

Соответственно матрица сил связности:

$$M = \sum_{k=1}^{\infty} k^{-1} E^k = -\ln(I - E).$$

В данном случае длинные пути дают значительно больший вклад в метрическую связность, чем в предыдущем случае. В частности, необходимо $p_n(v_i, v_j) = n$ путей длины n, чтобы получить вклад, эквивалентный дуге. И, соответственно, $p_n(v_i, v_j) = n/k$ путей длины n, чтобы получить вклад, эквивалентный пути длиной k.

Нахождение метрической связности вершин может применяться для решения двух основных задач:

Частная задача. Нахождение вершин графа, наиболее сильно связанных с заданной вершиной. Более строго формулировка задачи сводится к поиску вершин v', метрическая связаность которых с данной вершиной v лежит в некотором интервале In:

$$w(v, v') \in In.$$

При данном подходе существенным является выбор интервала In. Проанализируем влияние выбора интервала на смысл решаемой задачи для весовой функции $g(k) = 1/\alpha^{k-1}$ $(0 < \alpha < 1)$.

Рассмотрим несколько вариантов интервалов In:

$In = [1, \infty)$ – позволяет выявить все вершины, связанные с данной вершиной дугами, либо эквивалентные случаи;

$In = (1, \infty)$ – позволяет выявить вершины, которые связаны с данной дугой и еще, как минимум, одним путем;
$In = [1 + \alpha, \infty)$ – вершины, связанные с v дугой и, как минимум, путем с длиной 2;
$In = [1 + \alpha^n, \infty)$ – вершины, связанные с v дугой и, как минимум, путем с длиной не более $n + 1$;
$In = [\alpha, 1)$ – вершины, не связанные дугой с v, но связанные путем длины 2;
$In = [\alpha^2, \alpha)$ – вершины, не связанные дугой с v, но связанные путем длины 3;
$In = [\alpha^n, \alpha^{n-k})$ – вершины, не связанные дугой с v, но связанные путем длины не менее k и не более $n + 1$;

В этом случае k путей длины k эквивалентны дуге.

Общая задача. Разбиение графа на связные структуры, то есть поиск подграфов с определенной метрической связностью вершин. Как и для частной задачи, рассмотрим случаи различных весовых функций.

Необходимо ввести некоторый интервал значений метрической связности In для вершин, входящих в одну связную структуру. Более строго, связной структурой будем называть подмножество вершин $V' \subseteq V$, для каждой пары вершин из которого

$$w(v_i, v_j) \in In, \quad v_i, v_j \in V'.$$

Рассмотрим несколько вариантов интервалов In для весовой функции $g(k) = 1/\alpha^{k-1}$ $(0 < \alpha < 1)$:
$In = [1, \infty)$ – позволяет выявить связные структуры, все вершины которых попарно связаны друг с другом дугой, либо эквивалентные случаи;
$In = (1, \infty)$ – позволяет выявить связные структуры, все вершины которых попарно связаны друг с другом дугой и еще, как минимум, одним путем;
$In = [1 + \alpha, \infty)$ – позволяет выявить связные структуры, все вершины которых попарно связаны друг с другом дугой и, как минимум, путем с длиной 2;
$In = [1 + \alpha^n, \infty)$ – позволяет выявить связные структуры, все вершины которых попарно связаны друг с другом дугой и, как минимум, путем с длиной не более $n + 1$;
$In = [\alpha, 1)$ – позволяет выявить связные структуры, все вершины которых попарно не связаны друг с другом дугой, но связанные путем длины 2;
$In = [\alpha^2, \alpha)$ – позволяет выявить связные структуры, все вершины которых попарно не связаны друг с другом дугой, но связанные путем длины 3;
$In = [\alpha^n, \alpha^{n-k})$ – позволяет выявить связные структуры, все вершины которых попарно не связаны друг с другом, не связаны дугой с v, но связанные путем длины не менее k и не более $n + 1$;

2. Компьютерный эксперимент

Компьютерный эксперимент ставился с целью сравнения результатов, получаемых при различных весовых функциях $g(k)$. Исследовались ориентированные графы с количеством вершин равным 100 и некоторым определенным количеством дуг d. Матрица смешения задавалась случайным образом. Во всех экспериментах генерировалось 100 матриц, после чего производилось усреднение.

Матрица связности вычислялась с точностью до третьего слагаемого:

$$M^{(3)} = g(1)E + g(2)E^2 + g(3)E^3.$$

На рисунке 1 по оси абсцисс отложена величина метрической связности, а по оси ординат количество соответствующих связей.

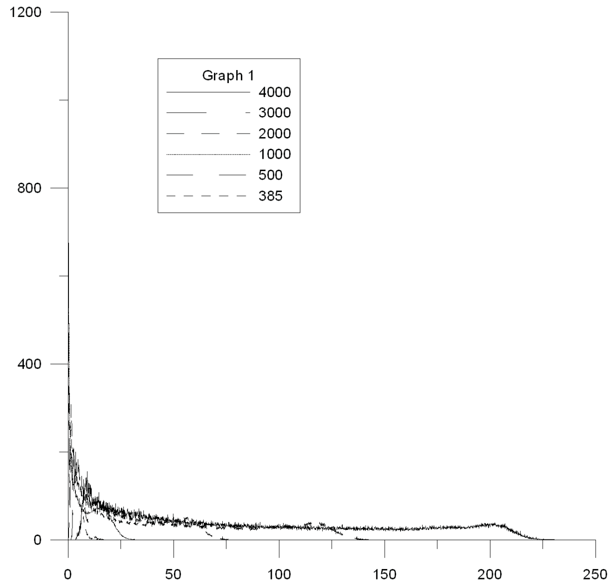

Рис. 1. Решение частной задачи для весовой функции $g(k) = \alpha^{k-1}$ $(\alpha = 0.1)$

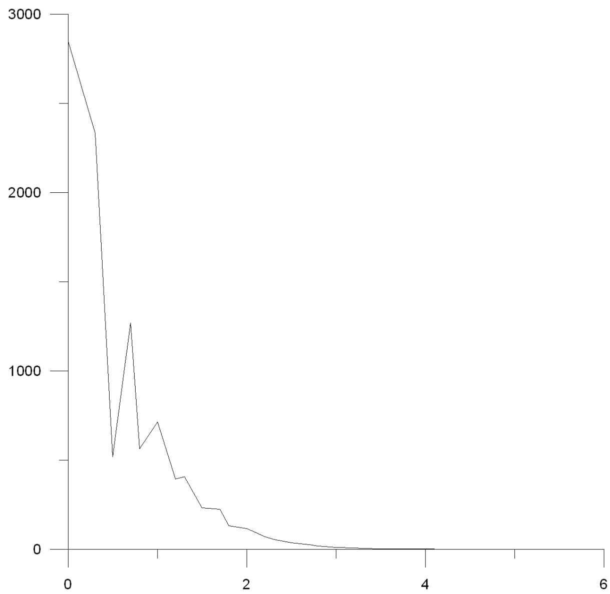

Рис. 2. Решение частной задачи для весовой функции $g(k) = 1/k$, $d = 500$

На рисунках 2 и 3 представлены аналогичные графики для весовой функции $g(k) = 1/k$ и $d = 500$(рис. 2) и $d = 3000$ (рис. 3).

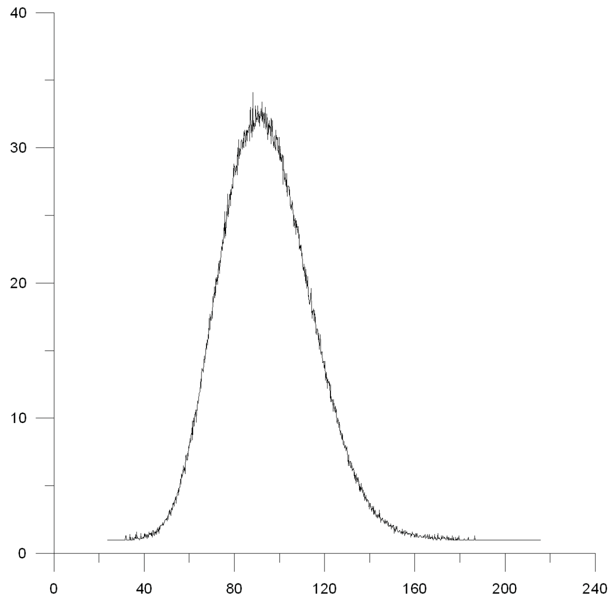

Рис. 3. Решение частной задачи для весовой функции $g(k) = 1/k$, $d = 3000$

Как видно, в случае весовой функции $g(k) = 1/k$ присутствует пик связностей определенной величины. Это обстоятельство делает данный вид весовой функции малопригодным для исследования графов в реальных задачах. Напротив, использование весовой функции $g(k) = \alpha^{k-1}$ приводит к адекватному поведению графиков, а именно: чем больше величина метрической связности, тем меньше соответствующее количество связей. Попытки получения точного результата по формуле

$$M = \alpha^{-1}((I - \alpha E)^{-1} - I)$$

наталкиваются на трудности, связанные с наличием ориентированных циклов, и, как следствие, с многократным учетом одних и тех же путей.

Авторы выражают благодарность Н.Ф. Богаченко за полезные обсуждения постановки задачи и используемой терминологии.

Литература

1. Girvan M., Newman M.E.J. Community structure in social and biological networks // arXiv:cond-mat/0112110v1 (2001).
2. Newman M.E.J. Mixing patterns in networks // Phys. Rev. E. 2003. V. 67. P. 026126-1–026126-13.