

УДК 519.178

АЛГОРИТМЫ ПОСТРОЕНИЯ ПРОСТРАНСТВЕННОГО СУПЕРГРАФА

Е.М. Сухова

Научный руководитель: доцент, к.ф.-м.н., М.Е. Семенов

Национальный исследовательский Томский политехнический университет,

Россия, г. Томск, пр. Ленина, 30, 634050

E-mail: ems16@tpu.ru

ALGORITHMS FOR CONSTRUCTING A SPATIAL SUPERGRAPH

E.M. Suhova

Scientific Supervisor: Ass. Pr., PhD, M.E. Semenov

Tomsk Polytechnic University, Russia, Tomsk, Lenin str., 30, 634050

E-mail: ems16@tpu.ru

***Abstract.** In the paper, we proposed two algorithms for constructing a supergraph in order to identify key relations in data with a network structure. On the Brightkite dataset, we have obtained the data compression by 3 and 9 times, respectively. The software implementation of the algorithms have been made using the NetworkX Python library.*

Введение. Граф представляет собой математическую абстракцию реальных систем любой природы, поэтому теория графов, как раздел математики, используется в самых разных предметных областях. Экспоненциальный рост объемов информации предопределяют необходимость разработки эффективных алгоритмов для преобразования исходных с сетевой структурой в графовые модели. Под *суперграфом* $SG(V', E')$ будем понимать подмножество вершин и ребер исходного графа $G(V, E)$, которые выявлены или заданы по определенному правилу, например, через функцию расстояния, учитывающую топологические отношения пространственных объектов [1, 2]. На таких моделях можно применять различные алгоритмы для решения задач кластеризации, выявления мостов, узлов сочленений, поиска покрытий, в том числе кликовых, имеющих прикладное значение в различных предметных областях.

Цель работы – создание и программная реализация алгоритма построения пространственного суперграфа для выявления ключевых отношений в данных, имеющих сетевую структуру.

Материалы и методы исследования. Для вычислительных экспериментов мы использовали набор данных с сетевой структурой Brightkite [3]. Данный набор содержит $m=214\,078$ рёбер и $n=58\,228$ вершин, а также атрибутивную информацию вершин: идентификатор пользователя (id), время регистрации в сети (check-in-time), координаты местонахождения пользователя (latitude – широта, longitude – долгота). Анализ исходных данных показал, что граф содержит 547 компонент связности, средняя степень вершин равна 7,353. В работе мы использовали максимальную компоненту связности с 56739 вершинами.

Алгоритм 1 (пространственный порог). Каждой вершине были присвоены декартовы координаты местонахождения пользователя, для всех ребер определен вес, как евклидово расстояние.

Далее для построения суперграфа мы ввели пространственный порог $R = \frac{1}{m} \sum_{i=1}^m p_i$, где p_i – длина ребра, m – количество ребер. Если вес ребра графа больше R , то ребро исходного графа удаляем, при этом это ребро запоминаем для дальнейшего использования (рис. 1, красные ребра). В итоге, мы получаем группу географически близко находящихся друг к другу вершин, которую будем называть *компонентой связности* (рис. 1, центр). Каждую компоненту стягиваем в одну вершину – эта вершина и будет вершиной суперграфа. Координаты супервершины находим как географический центр компоненты связности. Соединяем компоненты связности учитывая ранее помеченные ребра.

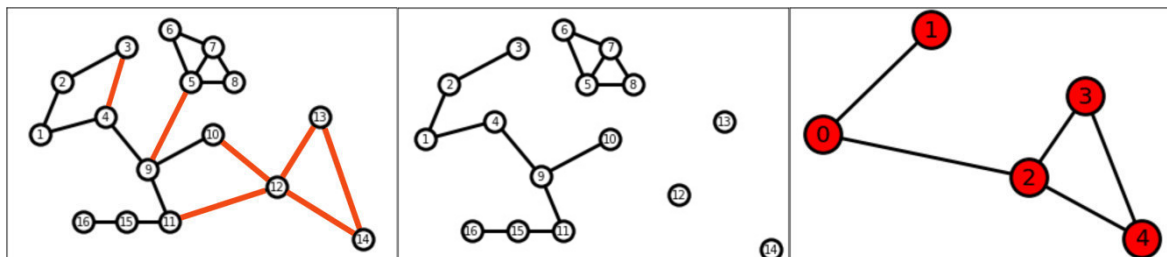


Рис. 1. Исходный граф, красным отмечены ребра с весом больше R (слева), расположение компонент связности (по центру), суперграф (справа)

Алгоритм 2 (мосты). Для проведения процедуры стягивания мы предлагаем использовать мосты (рис. 2, слева выделены красным). После удаления мостов получаем компоненты связности, которые затем стягиваем в одну вершину. Если в графе встречается цепь из двух и более мостов подряд, например, ребра (11, 15) и (15, 16), то мы удаляем только последнее ребро (11, 15), которое соединено с другой компонентой связности.

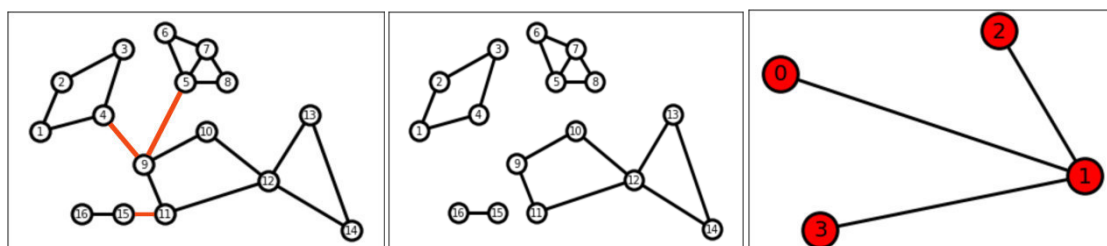


Рис. 2. Исходный граф, красным отмечены ребра-мосты (слева), расположение компонент связности (по центру), суперграф (справа)

Таким образом, Алгоритм 1 позволяет получить граф, в котором возможен цикл, а Алгоритм 2 позволяет получить дерево.

Результаты. С использованием пространственные характеристик (θ, φ) мы отобразили вершины графа на сферу (рис. 3 слева) в декартовой системе координат по формулам:

$$x = r \sin \theta \cos \varphi, y = r \sin \theta \sin \varphi, z = r \cos \theta,$$

где θ, φ координаты широты и долготы соответственно, $r = 6371$ км – радиус Земли. Для определения расстояние между всеми вершинами u и v мы использовали евклидово расстояние:

$$d = \sqrt{(x_i - x_j)^2 + (y_i - y_j)^2 + (z_i - z_j)^2}, i \neq j.$$

Зависимость количества вершин в суперграфе от порога R приведена на рис. 3. Видно, что при $R=100$ км количество вершин 28276, которое плавно снижается до 24113 с увеличением до $R=1000$ км, а затем претерпевает качественное изменение.

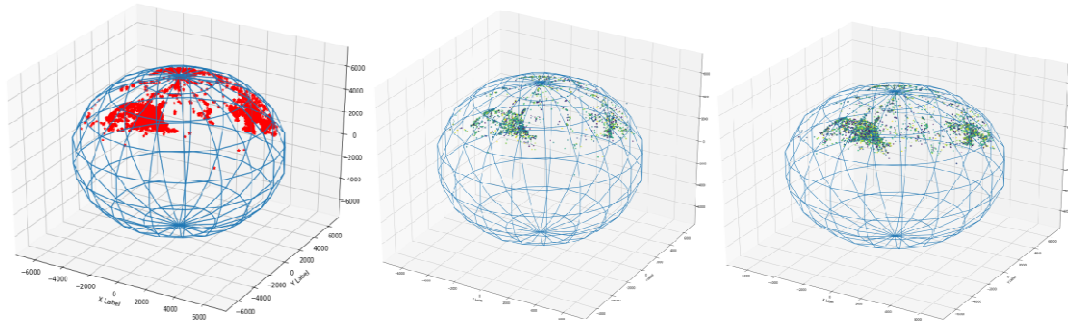


Рис. 3 Исходный граф (слева), результат работы 1 алгоритма (по центру), результат работы 2 алгоритма (справа)

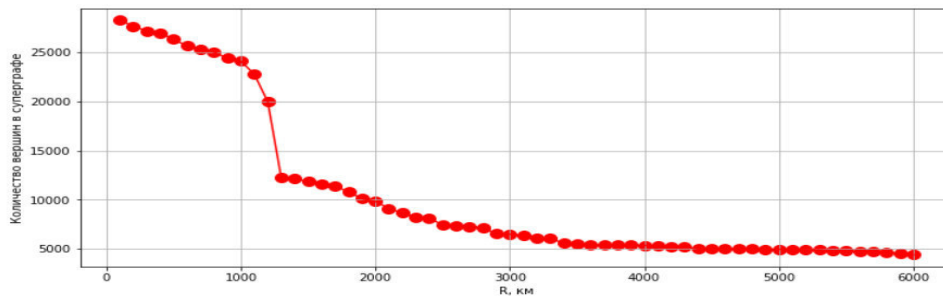


Рис. 4. Количество вершин графа в зависимости от пространственного порога

Заключение. Предложенные алгоритмы имеют практическое значение для выявления сообществ [4]. Таким образом, было разработано и программно реализовано два алгоритма построения суперграфа с использованием атрибутов вершин исходного графа. Алгоритм 1, на основе среднееарифметического расстояния, позволил нам сократить количество вершин в 9 раз, в то время как Алгоритм 2 с мостами позволил сократить в 3 раза. Использование данных алгоритмов значительно упрощает работу с сетевыми структурами.

СПИСОК ЛИТЕРАТУРЫ

1. Mukherjee S.S., Chakrabarti S. Graphon Estimation from Partially Observed Network Data. 2019., p. 1-12
2. Еремеев С.В. Алгоритм построения моделей пространственных отношений на основе темпоральных графов // Сборник трудов III международной конференции и молодежной школы «Информационные технологии и нанотехнологии» – Самара: Новая техника, 2017. – С. 482-489.
3. Cho E., Myers S. A., Leskovec J.. Friendship and Mobility: Friendship and Mobility: User Movement in Location-Based Social Networks ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD), 2011. <https://snap.stanford.edu/data/loc-brightkite.html>
4. Scripps J., Tan P., Esfahanian A. Node Roles and Community Structure in Networks. New York: ACM, 2007. 36 p.