Филиал Московского Государственного Университета

имени М.В.Ломоносова в городе Ташкенте

Факультет прикладной математики и информатики

Кафедра прикладной математики и информатики

**Борисов Артур Александрович**

**ВЫПУСКНАЯ КВАЛИФИКАЦИОННАЯ РАБОТА**

**на тему: «О нахождении степени посредничества в динамически меняющихся структурах»**

**по направлению 01.03.02 «Прикладная математика и информатика»**

ВКР рассмотрена и рекомендована к защите зав. кафедрой «ПМиИ», д.ф.-м.н., профессором, Гасановым Э. Э.

Научный руководитель: к.ф.-м.н. **Алексеев Дмитрий Владимирович**

«20» мая 2024 г.

Ташкент 2024

Аннотация

В данной работе рассматривается задача нахождения «Степени посредничества» для неориентированных невзвешенных графов. Цель работы заключается в разработке алгоритма определения данной степени на основе Алгоритма Брандеса[1] c динамическим пересчетом и допустимой погрешностью; сравнение с существующими методами из библиотеки Networkx[2]; и реализация интерфейса с динамическим изменением структуры и визуализации вычисляемых метрик.

**Abstract**

In this paper, we consider the problem of finding "Betweenness centrality" for undirected unweighted graphs. The aim of the work is to develop an algorithm for determining the “Betweenness centrality” based on the Brandes Algorithm [1] with dynamic recalculation and some deviation, comparison with existing methods from the Networkx library [2]; and implementation of an interface with a dynamic change in the structure and visualization of calculated metrics.

Оглавление

[Аннотация 2](#_Toc167738922)

[Введение 4](#_Toc167738923)

[Постановка задачи 5](#_Toc167738924)

[Готовые решения 7](#_Toc167738925)

[Динамический алгоритм пересчета степени посредничества 8](#_Toc167738926)

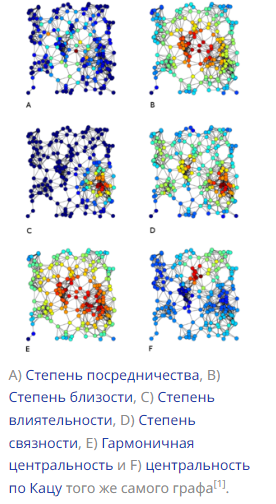
[Интерфейс разработанной программы 9](#_Toc167738927)

[Тестирование 10](#_Toc167738928)

[**Заключение** 11](#_Toc167738929)

[Список использованных источников и литературы 12](#_Toc167738930)

[Приложения 13](#_Toc167738931)

Введение

Понятие степени посредничества (betweenness centrality) в теории графов было впервые введено в 1977 году социологом Линусом Фрименом. Он использовал это понятие для изучения взаимодействия между участниками в социальной сети.

Одной из задач, которая решается с помощью степени посредничества, является поиск важных узлов в сетях. Например, в социальной сети важные узлы могут быть связаны с влиятельными людьми или общими интересами. В биологической сети важные узлы могут означать гены, ответственные за регуляцию других генов.

Другой задачей, связанной со степенью посредничества, является поиск наиболее быстрых маршрутов в сетях. Как правило, маршруты через узлы с высоким посредничеством более быстрые и надежные.

Алгоритмы точного вычисления Betweenness Centrality существуют и основаны на переборе всех кратчайших путей в графе. Однако такие алгоритмы становятся чрезвычайно времязатратными, особенно при работе с большими графами, где требуется добавить или удалить небольшое количество ребер или вершин.

Постановка задачи

**Определение 1.** Неориентированным графом 𝐺 называется пара 𝐺 = (𝑉, 𝐸), где 𝑉 — конечное непустое множество, а 𝐸 ⊂ 𝑉, то есть является множеством неупорядоченных пар различных элементов из 𝑉. Элементы множества 𝑉 называются вершинами, а множества 𝐸 —– ребрами.

**Определение 2.** Граф G, называется ориентированным, если 𝐸 ⊂ 𝑉 × 𝑉, то есть множество 𝐸 является подмножеством упорядоченных пар различных элементов из 𝑉. В таком случае элементы множества 𝐸 также могут называться дугами.

**Определение 3**. Граф 𝐺 = (𝑉, 𝐸) называется взвешенным, если его каждому из его ребер поставлено в соответствие некоторое числовое значение (вес ребра).

**Определение 4**. Путем в графе называется последовательность вида

где ∈ 𝑉, ∈ 𝐸, = ( , ), 𝑖 = 1, …, 𝑘.

**Определение 5.** Дана весовая функция , которая отображает ребра на их веса, значения которых выражаются действительными числами, и неориентированный граф G. Тогда **кратчайшим путём** из вершины в вершину будет называться путь (где и ), который имеет минимальное значение суммы ). Если все ребра в графе имеют единичный вес, то задача сводится к определению наименьшего количества обходимых ребер.

**Определение 6.** Пусть

Обозначим:

– кратчайший путь между вершинами «s» и «t» в графе G

– кол-во кратчайших путей из «s» в «t»

– кол-во кратчайших путей из «s» в «t» в которых есть вершина «v»

– множество предшественников вершины «v» на кратчайших путях из «s»

**Определение 7.** Степень посредничества — это мера центральности в графе, основанная на кратчайших путях. Для любой пары вершин в связном графе существует по меньшей мере один (кратчайший) путь между вершинами, для которого минимально либо число рёбер, по которым путь проходит, (для невзвешенных графов), либо сумма весов этих рёбер (для взвешенных графов). Степень посредничества для каждой вершины равна числу этих кратчайших путей через вершину.

При вычислении степеней посредничества всех вершин в графе требуется вычисление кратчайших путей между всеми парами вершин графа.

На невзвешенных графах вычисление степени посредничества с помощью алгоритма Брандеса занимает время

Готовые решения

Библиотека Networkx [2].

Модуль позволяет:

* строить различные типы графов и реализовывать сложные алгоритмы сетевого анализа.
* хранить базы данных, оформленные в виде сетей как в собственных форматах данных, так и в обычном графическом виде.

Достоинствами данного модуля является простота освоения при обладании широкой сферой использования.

В библиотеке NetworkX для нахождения степени посредничества графа G можно использовать следующие метод на основе алгоритма Ульрика Брандеса [1]:

***betweenness\_centrality(G, normalized, weight, endpoints, seed)***

* Параметр normalized (по умолчанию False) определяет, должны ли значения степени посредничества быть нормализованы в диапазоне от 0 до 1.
* Параметр weight позволяет учитывать веса ребер при вычислении степени посредничества.
* Параметр endpoints указывает, должны ли конечные вершины ребер учитываться при вычислении степени посредничества.
* Параметр seed задает начальное значение для генератора случайных чисел.

Оценка времени выполнения равна *O*(n\*m) в невзвешенных сетях и в взвешенных, n – количество вершин, m — количество ребер) при использовании O(n+m) памяти.

Динамический алгоритм пересчета степени посредничества

В ходе работы Алгоритм Брандеса был модернизирован возможностью динамического пересчета «Степени посредничества»,с некоторой ошибкой при добавлении/удалении вершин/ребер в граф с уже точно подсчитанной степенью.

def brandes\_partial(G, affected\_nodes)

G – граф, для вершин которого требуется рассчитать степень

affected nodes – подграф Ĝ

Способ выделение в графе G подграфа Ĝ:

* в случае добавления вершин/ребер Ĝ состоит из добавленных вершин и их соседей в 2 поколениях;
* в случае удаления вершин/ребер, Ĝ состоит только из советующих соседей затронутых вершин в 2 поколениях;

**Оценка времени** перерасчета равна *O*(ñ\*ɱ), ñ – количество вершин в Ĝ, ɱ — количество ребер в Ĝ.

**Средняя ошибка отклонения** при таком пересчете в большей мере зависит от места добавления вершины, нежели от размерности графа:

Пример 1. Дан граф G = nx.karate\_club\_graph() состоящий из 33 вершин и 78 ребер при добавлении вершины 34 и ребер (34, 7), (34, 1) общий процент ошибки составляет **32.29%** (вершины 1 и 7 при первом полном вычислении имели минимальную степень [рис.1 в приложении]), а при добавлении ребер (34, 0), (34, 2) = **8.11%** (вершины 0 и 2 при первом полном имели максимальную степень [рис.2]).

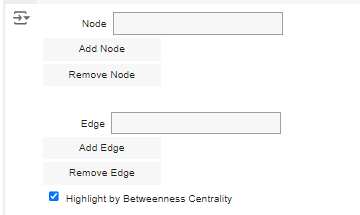
Вывод: чем ближе происходит добавление вершин/ребер к наиболее важным участкам, тем меньше среднее отклонение.

Пример 2. Дан граф G состоящий из 599 вершин из 5304 ребер при добавлении 600 вершины и ребер (600, 1), (600, 565) общий процент ошибки составил **26.51%**

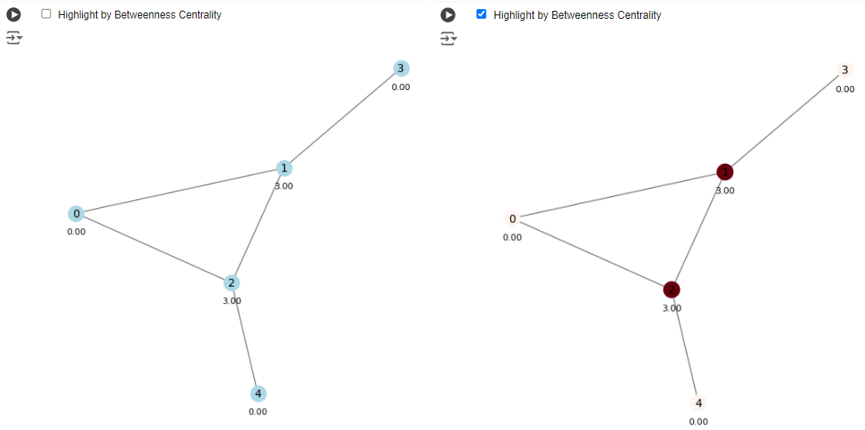
Вывод: Увеличение размерности графа сильно не влияет на изменение среднего отклонения, при добавлении малого числа вершин/ребер.

Интерфейс разработанной программы

Реализация интерфейса выполнена в бесплатной интерактивной облачной среде для работы с кодом на языке Python от Google – Colab.



Первое поле отвечает за добавление вершин, Второе за добавление ребер. Галочка в нажатом состоянии красит градиентом вершины в соответствии с их степенью (в не нажатом графа не раскрашен).



Под каждой вершиной указывается полученное значение степени посреднечества.

Тестирование

Производилось тестирование программы на датасетах:

* nx.les\_miserables\_graph - Совместного появления персонажей в романе В.Гюго "Отверженные" [5] (рис 3).
* nx.florentine\_families\_graph – связь флорентийских семей [6] (рис 4).
* Матчи по американскому футболу между колледжами дивизиона IA в течение регулярного сезона осенью 2000 года, составленную М. Гирваном и М. Ньюманом [7] (рис 5).

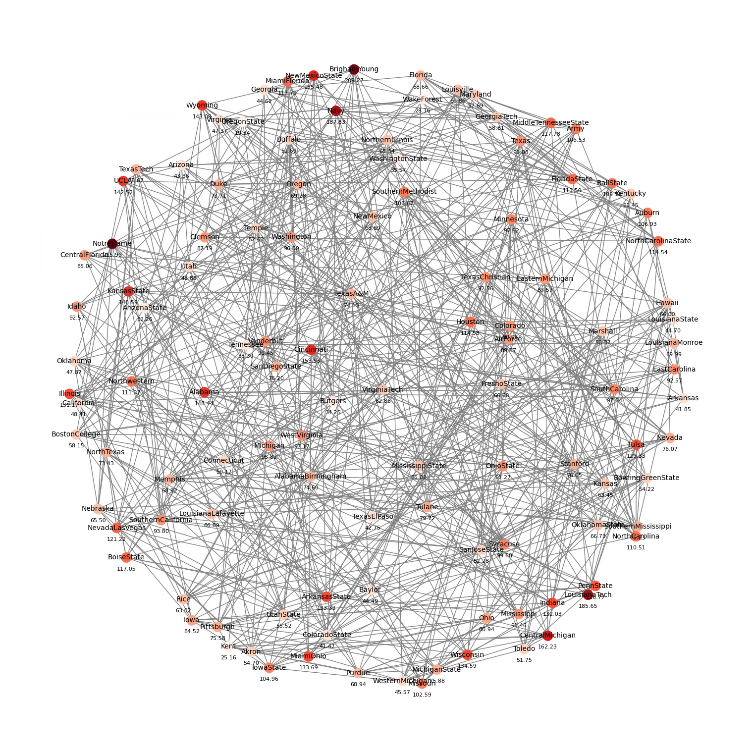


Рис.5

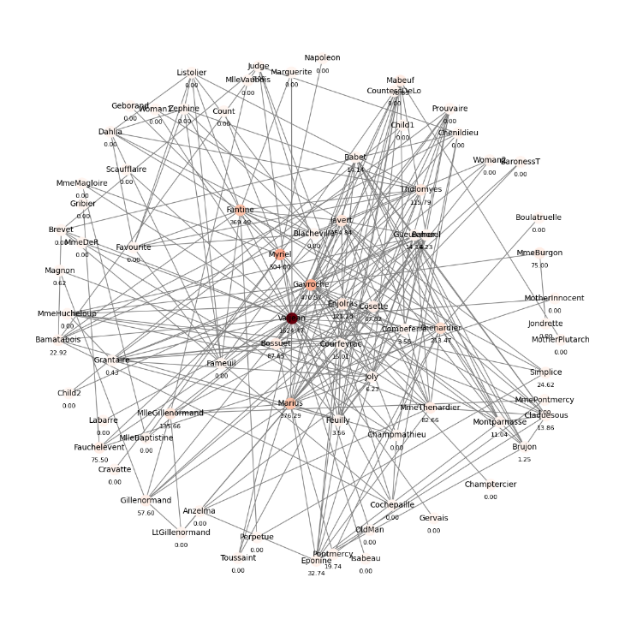


Рис. 3

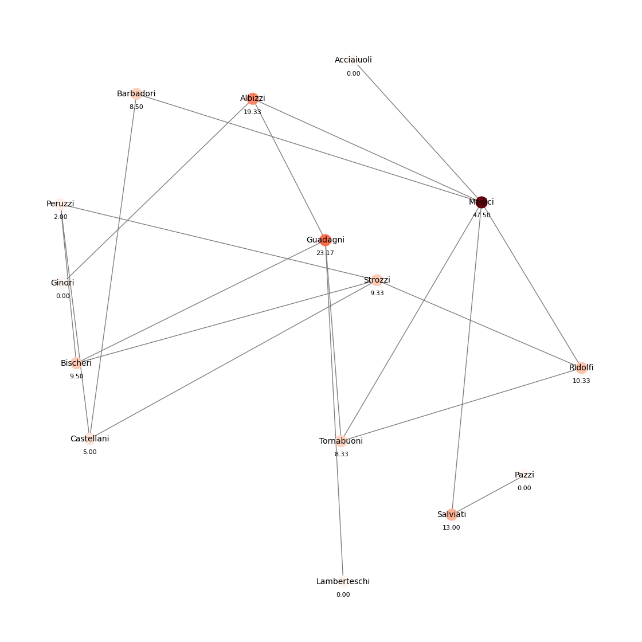


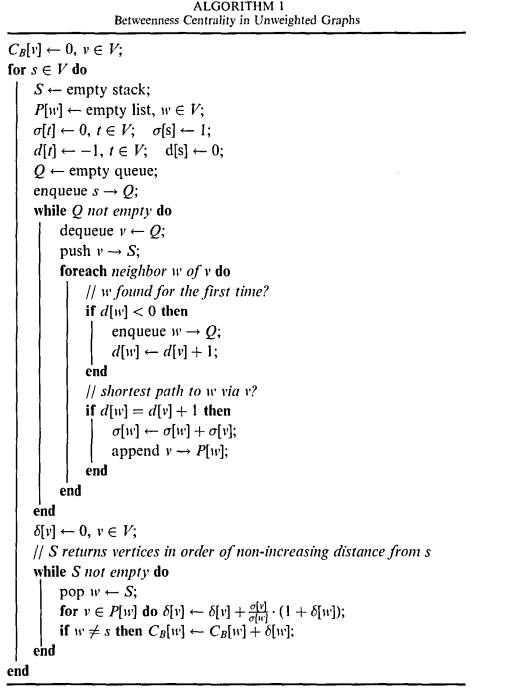
Рис 4

# Заключение

В рамках дипломной работы был реализован алгоритм Брандеса для нахождения степени посредничества в неориентированных невзвешенных графах при динамическом пересчете с допустимой погрешностью. Разработана программа динамического изменения графа с пересчетом значений степеней, и визуализация с выделением вершин градиентом по возрастанию значения степени. Все результаты выложены в открытый доступ на GitHub. Ссылка на репозитарий: <https://github.com/ArturBorisov01/Algoritm-Brandes/tree/master>

Список использованных источников и литературы

1. Статья на Википедии. Определение и описание индексов центральности <https://w.wiki/6kaC>
2. Документация библиотеки Networkx для python <https://networkx.org/documentation/stable/tutorial.html>
3. Ульрик Брандес: более быстрый алгоритм централизации промежуточности . Журнал математической социологии 25 (2): 163–177, 2001 г. [https://doi.org/10.1080/0022250X.2001.9990249 .](https://doi.org/10.1080/0022250X.2001.9990249%20.)
4. Веб среда разработки <https://colab.research.google.com/>
5. D. E. Knuth, 1993. The Stanford GraphBase: a platform for combinatorial computing, pp. 74-87. New York: AcM Press.
6. Ronald L. Breiger and Philippa E. Pattison Cumulated social roles: The duality of persons and their algebras,1 Social Networks, Volume 8, Issue 3, September 1986, Pages 215-256
7. Дата сет Американский футбол <https://networkx.org/documentation/stable/auto_examples/graph/plot_football.html>

Приложения

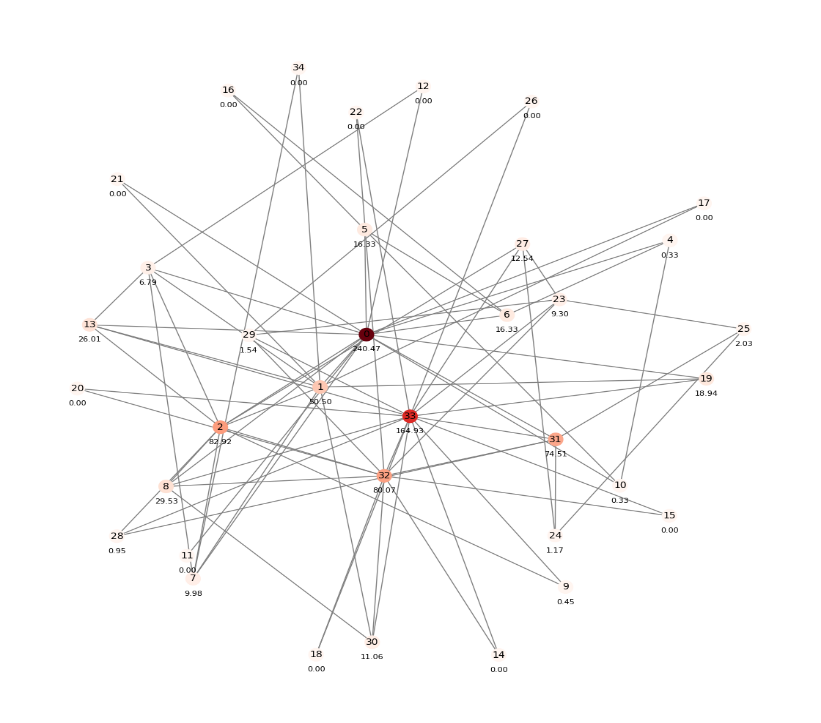


Рис. 2

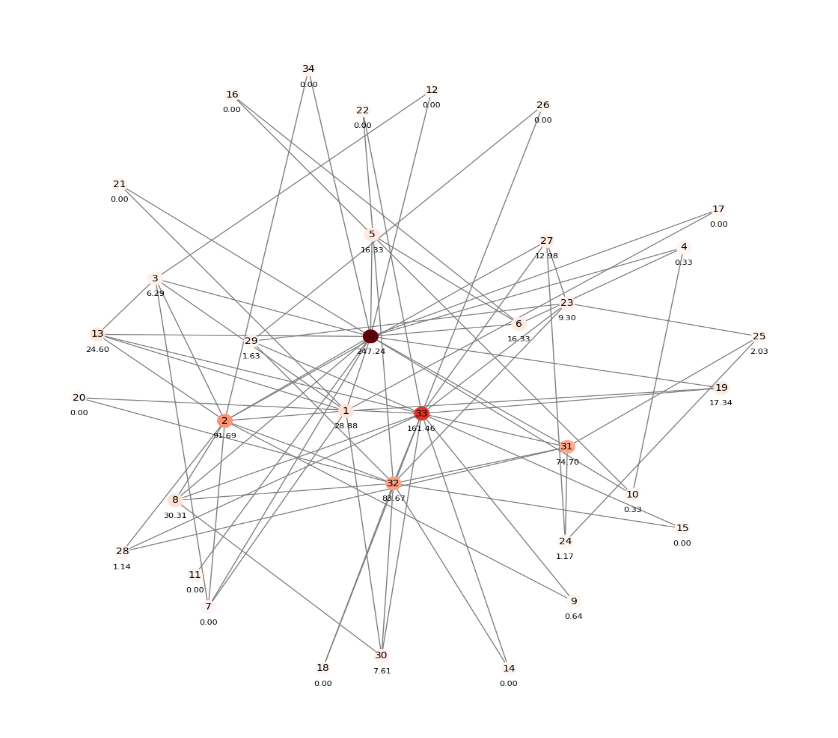


Рис. 1