

Забайкальский государственный университет

На правах рукописи

Цынгуев Булат Тимурович

**МАТЕМАТИЧЕСКИЕ МОДЕЛИ РАНЖИРОВАНИЯ ВЕРШИН
В ГРАФАХ КОММУНИКАЦИОННЫХ СЕТЕЙ**

05.13.18 — «Математическое моделирование, численные методы и комплексы программ»

ДИССЕРТАЦИЯ

на соискание ученой степени
кандидата технических наук

Научный руководитель:
доктор физико-математических наук, профессор
В. В. Мазалов

Чита – 2015

Оглавление

Введение	4
Глава 1. Электрическая центральность вершин во взвешенных графах социальных сетей	14
1.1. Классическая центральность	14
1.2. PageRank	18
1.3. Центральность на основе правил Кирхгофа	27
1.4. Мера близости в модели электрических цепей	32
1.5. Центральность вершин звезды	33
1.5.1. Невзвешенная звезда	33
1.5.2. Звезда с одним взвешенным ребром с весом равным k	34
1.6. Центральность вершин полного двудольного графа	36
1.6.1. Двудольный граф $K_{2,n-2}$	37
1.6.2. Двудольный граф $K_{3,n-3}$	38
1.6.3. Двудольный граф $K_{r,n-r}$	40
1.7. Сравнение с PageRank и классической центральностью	42
1.7.1. Простой пример взвешенного графа	42
1.7.2. Пример взвешенного графа с 6 вершинами	43
1.7.3. Электрическая центральность двух звезд	44
1.8. Приложение для социальных сетей	47
Глава 2. Теоретико-игровые модели центральности вершин в коммуни- кационных графах	55
2.1. Определение характеристической функции в кооперативной игре с помощью электрической центральности	55
2.2. Использование вектора Майерсона для ранжирования вершин ком- муникационного графа	58
2.3. Сравнение с другими моделями ранжирования	62

Глава 3. Компьютерное моделирование в задачах анализа коммуникационных сетей	66
3.1. Приложения для транспортных сетей	66
3.1.1. Транссибирская железнодорожная магистраль	66
3.1.2. Железные дороги Финляндии	70
3.1.3. Железные дороги Китая	71
3.1.4. Метро города Москвы	73
3.1.5. Граф стран Евразии	76
3.2. Анализ портала математических публикаций Math-Net.ru	77
3.3. Ранжирование сайтов научных организаций	81
3.4. Онтологическая модель	86
Заключение	93
Литература	95
Приложение А. Листинги программного комплекса	103
Приложение Б. Копия свидетельства о регистрации программы	109

Введение

Актуальность темы.

Методы анализа социальных сетей (*social network analysis*) применяются во многих областях науки, таких как экономика, физика, социология, биология и информационные технологии. Важность роли сетевых структур в социологических исследованиях на сегодняшний день не подвергается сомнению. Знаменитый пример, исследованный в [31, 56, 68] просто и наглядно объясняет могущество, и власть семьи Медичи на примере сети брачных отношений между ведущими семьями Флоренции начала 15 века. Существуют также множество примеров того, как сетевые структуры взаимоотношений между людьми позволяют выявлять и объяснять новые и более глубокие знания об исследуемых социальных сетях. К ним можно привести такие работы как [30, 39, 79].

Информационные технологии также является традиционной областью применения методов ранжирования, например, широко известный метод ссылочного ранжирования PageRank [34]. В основе любых современных поисковых систем лежат механизмы ранжирования, например, для крупнейшей поисковой системы Google для ранжирования веб-страниц используется метод PageRank [35]. В работе [57] приведены примеры того, что использование методов структурного ранжирования участников одноранговых сетей (peer-to-peer) позволяет улучшить базовые показатели работы всего протокола обмена данными. В [78] для ранжирования участников одноранговых сетей в протоколе обмена данными применяется распределенный PageRank.

Со временем появилась необходимость в анализе более сложных сетевых структур так называемых сложных сетей (complex networks). Здесь можно отметить работы [26, 64, 73, 77].

Современные методы исследования сетевых структур получили бурное развитие в связи с появлением феномена социальных сетей.

Одним из базовых понятий в анализе сетевых структур является централь-

ность (betweenness centrality) [43]. Первое описание центральности было в [42]. Центральность вершины v - это важная мера, отражающая то, насколько вершина v участвует в процессе распространения информации между остальными вершинами в графе.

Одно из первых определений центральности представлено следующей формулой:

$$c_B(v) = \frac{1}{n_B} \sum_{s,t \in V} \frac{\sigma_{s,t}(v)}{\sigma_{s,t}}, \quad (1)$$

где $\sigma_{s,t}$ - число геодезических путей между вершинами s и t , $\sigma_{s,t}(v)$ - число геодезических путей между вершинами s и t проходящих через вершину v . Коэффициент нормировки n_B равен $n_B = (n-1)(n-2)$, если вершина v не может быть начальной s или конечной t вершинами, и $n_B = n(n-1)$ иначе. Для невзвешенного графа алгоритм поиска центральности с наилучшей вычислительной сложностью представлена в [28], вычислительная сложность алгоритма равна $O(mn)$.

Одним из основных недостатков данной меры центральности является то, что в расчетах учитываются только геодезические пути, и не берутся в расчет пути, которые могут быть длиннее геодезических всего на один или два шага, хотя данные негеодезические пути могут играть важную роль в описании и объяснении процессов протекающих в коммуникационных сетях. К тому же в [45, 63] были представлены доказательства того, что меры центральности вершин (1) сильно коррелированы со степенью вершин. Вскоре У. Брандесом и Д. Флейсчером [29] и М. Ньюманом [65] были предложены способы расчета центральностей, учитывающие негеодезические пути. Данные модели получили название электрическая центральность (current flow betweenness centrality) или другими словами, мера центральности на основе модели электрических цепей. В [29, 65] граф рассматривается как электрическая цепь с идеальными элементами, где каждое ребро имеет некую пропускную способность (значение обрат-

ное сопротивлению), а вершины графа являются ее узлами. Для поиска меры центральности в модели электрической цепи используются правила Кирхгофа. Для этого электрическая цепь заземляется в некоторой вершине t и подается электрический ток в некоторой вершине s . В [29] ток подается в некоторой единственной вершине s , а также в единственной вершине t сеть заземляется. Мерой центральности вершины v служит средняя величина тока, проходящего через вершину v по всем возможным парам s и t . Таким образом, в модели электрической цепи при расчете меры центральности учитываются не только геодезические пути.

Наилучшая вычислительная сложность известного на текущий момент алгоритма поиска меры центральности в модели электрической цепи [29, 65] равна $O(I(n-1) + mn \log n)$, где $I(n-1)$ - сложность вычисления обратной матрицы размерности $n-1$.

Таким образом, вычислительная сложность алгоритма поиска меры центральности в модели электрической цепи является достаточно высокой. В сравнении с оригинальной центральностью, основанной на геодезических путях, узким местом в вычислении меры электрической центральности является нахождение обратной матрицы с вычислительной сложностью $O(n^3)$. Чтобы улучшить ситуацию с вычислительной сложностью алгоритма поиска электрической меры центральности в [14] предлагается модификация в модель электрической цепи, где в дополнение к заземленной вершине, каждая вершина соединена с заземлением с некоторой малой электропроводностью, пропорциональной степени этой вершины. Эти изменения позволили свести линейную систему уравнений к строгому диагональному виду, и существенно уменьшили вычислительную сложность, но по-прежнему требовали вычисления средних значений по всем возможным s и t . Отметим также, что данная модель применима только для невзвешенных графов.

В дополнение ко всему оригинальная центральность, основанная на геодезических путях, может быть легко расширена для ранжирования вершин графа

с целочисленными весами, например, путем замены каждого ребра веса k на k параллельных ребер весом в 1. Таким образом, задача сводится к исследованию мультиграфа. Геодезический путь между двумя вершинами определяется, так же как и для невзвешенных графов, но число геодезических путей становится больше за счет увеличения числа ребер. Допустим, что если вершины i_1 и i_2 соединены k ребрами, а вершины i_2 и i_3 соединены l ребрами, тогда вершины i_1 и i_3 соединены $k \cdot l$ путями. Применяя формулу (1) к вершинам мультиграфа, можно получить значение центральности для взвешенного графа, но со значительным увеличением вычислительной сложности. В наихудшем случае, когда имеется k ребер между любыми двумя вершинами вычислительная сложность центральности составит $O(mn^k)$.

Для определения центральности вершин графа можно также использовать методы кооперативной теории игр на основе модели *communication game*, предложенной Р. Майерсоном. Ситуация, в которой любое подмножество множества N игроков может сформировать коалицию и сотрудничать, получая определенные выплаты, может быть описана кооперативной игрой. Однако, существует много реальных ситуаций, для которых требуется модель, принимающая во внимание ограничения в сотрудничестве из-за различных политических, экономических, технических и даже психологических причин (антимонопольное законодательство, отсутствие каналов связи, личные симпатии и антипатии и т.д.). В известной статье Р. Майерсона [62] ограниченная кооперация представлена неориентированным коммуникационным графом. Вершины графа представляют игроков, а рёбра представляют связи между игроками. Игроки могут взаимодействовать напрямую, только если они связаны. Понятие связи при этом может интерпретироваться весьма широко: ее наличие может обозначать передачу информации или ресурсов между игроками, отношения сотрудничества и дружбы, наличие транспортной связи. Вершины могут быть отдельными людьми, организациями, странами или веб-страницами. Это приводит к так называемой коммуникационной игре, заданной тройкой, состоящей из конечного

множества игроков, характеристической функции и графа. Для такого класса игр Р. Майерсон определил и охарактеризовал правило распределения, являющееся обобщением вектора Шепли [71]. Впоследствии это правило было названо вектором Майерсона [13].

Также отметим, что традиционно в кооперативной теории игр для определения силы игроков используются индексы Шепли-Шубика [72], Банцафа [19], Дигана-Пакела [40] и Холера [47]. Данные индексы широко применяются в различных областях жизни, в частности в определении силы влияния партий в парламенте государства. Но указанные индексы не используются в случаях, когда возможность кооперации игроков задается сетевой структурой.

В 1996 Джексон и Волински [50] предложили модель, названную сетевой игрой, в которой экономические возможности зависят непосредственно от структуры сетей, соединяющих игроков и обобщили вектор Майерсона в контексте сетевых игр. Несмотря на то, что было предложено и изучено много других правил распределения (см., например, [24, 25, 36, 49, 69, 70, 74]), вектор Майерсона широко используется в качестве схемы выплат в сетевых играх.

Известно, что вычисление вектора Майерсона требует больших усилий. В работе [61] для игры со специальной характеристической функцией была предложена процедура получения дележа, в которой используется производящая функция, и доказано, что полученный в результате делёж совпадает с вектором Майерсона.

В России в области сетевых игр работают К. Е. Авраченко, А. Ю. Гарнаев, Д. А. Губанов, Д. А. Новиков, М. В. Губко, А. Г. Чхартишвили.

В данной работе предлагается новый способ определения центральности для взвешенных графов в модели электрических цепей. В данной модели электрической цепи каждая вершина цепи соединена с искусственно вводимой вершиной v_{n+1} ребром с пропускной способностью δ . Электрический ток подается в некоторую вершину s , цепь каждый раз заземляется в одной и той же вершине v_{n+1} . Таким образом, электрический ток течет по цепи из вершины s в

вершину v_{n+1} . Мерой центральности вершины v служит средняя величина тока, проходящего через вершину v по всем возможным s . Отметим, что ток на ребрах инцидентных с v_{n+1} вершиной не учитывается при расчете центральности. Рассматриваемый способ имеет относительно невысокую вычислительную сложность $O(n^3)$ равную сложности вычисления обратной матрицы. В отличие от оригинальной центральности вычислительная сложность алгоритма поиска предлагаемой электрической центральности, остается постоянной, как в случае с взвешенным графом, так и для невзвешенного графа.

Также в данной работе представлена модель с использованием кооперативной теории игр, где вершины графа рассматриваются в качестве игроков, а характеристическая функция определяется как количество проходящего через узел электрического тока в рассматриваемой нами модели электрических цепей. Далее центральность вершин определяется как решение кооперативной игры, т.е. вектором Майерсона. На ряде примеров сделано сравнение предлагаемого способа определения центральности с широко используемым методом ссылочного ранжирования PageRank.

Цель диссертационной работы заключается в построении и исследовании свойств математических моделей ранжирования вершин в графах коммуникационных сетей. Достижение поставленной цели требует решения следующих **задач**:

1. Повышение эффективности алгоритма поиска меры центральности вершин графа на основе модели электрических цепей. В качестве критерия эффективности принимается вычислительная сложность алгоритма;
2. Исследование центральности вершин графа в модели электрических цепей с использованием кооперативной теории игр;
3. Создание программного комплекса для численной реализации на ЭВМ исследуемых алгоритмов и проведение численных экспериментов.

Методы исследования. В диссертационной работе применяются методы теории графов, линейной алгебры и кооперативной теории игр. Для численного

моделирования и программной реализации разработанных алгоритмов применяются методы вычислительной математики и прикладного программирования.

Научная новизна работы заключается в следующем:

1. Предложен новый метод вычисления электрической центральности вершин, который позволяет уменьшить вычислительную сложность алгоритма поиска центральности вершин взвешенного графа.
2. Предложен алгоритм приближенного вычисления электрической центральности методом последовательных приближений для ранжирования вершин в графах большой размерности.
3. С использованием методов кооперативной теории игр построена модель, характеристическая функция которой строится на основе электрической центральности вершин. С помощью программной реализации на ЭВМ проведены численные эксперименты.

Практическая ценность работы заключается в реализации программы вычисления центральности вершин графа на основе правил Кирхгофа. Данная программа позволяет определить меру центральности вершин в графе путем определения величины среднего тока, проходящего через данную вершину в модели электрической цепи. В программе предусмотрена возможность ранжирования вершин графа на основе правил Кирхгофа с использованием метода последовательных приближений. Программа может быть использована для анализа социальных сетей, коммуникационных сетей и веб-сайтов. Программа прошла государственную регистрацию в Федеральной службе по интеллектуальной собственности, патентам и товарным знакам.

Положения, выносимые на защиту. На защиту выносятся следующие положения:

1. Предложен и исследован новый метод ранжирования вершин взвешенного графа на основе правил Кирхгофа. Предлагаемый метод сохранил полезные свойства методов ранжирования на основе моделей электрических цепей, но при этом имеет преимущества в вычислительной сложности.

2. Предложен способ приближенного вычисления мер центральностей для ранжирования вершин в графах большой размерности. Благодаря этому можно применять данный способ ранжирования на множестве прикладных задач.
3. Предложена и исследована модель с использованием методов кооперативной теории игр, характеристическая функция которой строится на основе электрической центральности.
4. Создан комплекс программ для численной реализации на ЭВМ предложенных алгоритмов и проведены численные эксперименты.

Связь работы с научными программами, темами: основные результаты диссертации были получены в рамках выполнения исследований при финансовой поддержке РГНФ (проект 15-02-00352) и Отделением математических наук РАН.

Апробация работы. Основные результаты работы были представлены и обсуждены на следующих семинарах и конференциях:

1. Международный семинар „Networking Games and Management“, Петрозаводск, Россия, 23-25 июня, 2013;
2. Международная конференция „Дифференциальные уравнения и математическое моделирование“, Улан-Удэ, Байкал, Россия, 22-27 июня, 2015;
3. Международный семинар „Networking Games and Management“, Петрозаводск, Россия, 5-7 июля, 2015;
4. IV Международная конференция по вычислительным социальным сетям (CSoNet 2015), Пекин, Китай, 4-6 августа, 2015;

а также на семинарах кафедры фундаментальной и прикладной математики, теории и методики обучения математике факультета естественных наук, математики и технологии Забайкальского государственного университета и на семинарах лаборатории математической кибернетики ИПМИ КарНЦ.

Публикации. Материалы диссертации опубликованы в 7 печатных работах, из них 2 статьи в рецензируемых журналах, рекомендованных ВАК [4, 9], 1

статья в издании, индексируемом в библиографической базе данных Scopus [18] и 3 тезисов докладов [6, 10, 11]. Получено свидетельство о государственной регистрации программы для ЭВМ [5].

Структура и объем работы. Диссертация состоит из введения, трех глав, разбитых на разделы и подразделы, заключения и списка литературы. Общий объем рукописи составляет 102 страницы. Работа содержит 22 рисунка и 30 таблиц.

Во **введении** отражена актуальность диссертационной работы, сформулирована цель исследования, обоснована научная новизна работы, показана практическая значимость полученных результатов, а также представлены положения, выносимые на защиту.

В **первой главе** рассматривается модель электрических цепей в задаче ранжирования вершин в графе. Приводятся основные определения и понятия. Представлен новый способ определения электрической центральности и меры близости. Предложен алгоритм приближенного вычисления электрической центральности методом последовательных приближений для ранжирования вершин в графах большой размерности. Найдены в аналитическом виде меры центральности вершин для некоторых частных видов графов: полный двудольный граф и частный случай звезды с одним взвешенным ребром. На ряде простых примеров проведен сравнительный анализ свойств электрической центральности с методом ранжирования PageRank, где представлены преимущества применения электрической центральности. Представлен пример использования электрической центральности для ранжирования графа сообщества „Теория игр“ в социальной сети ВКонтакте, также приведено сравнение результатов ранжирования с методом PageRank.

Во **второй главе** представлена теоретико-игровая модель для вычисления центральности вершин. В предлагаемой модели для ранжирования вершин используется известный метод кооперативной теории игр - вектор Майерсона. В данной модели вершины графа рассматриваются в качестве игроков, возмож-

ность кооперации игроков ограничена взвешенным графом. Вес ребра в графе отражает силу возможной кооперации соответствующих данному ребру игроков. Характеристическая функция строится на основе значений электрической центральности вершин. Предложенный метод ранжирования демонстрируется на некоторых известных примерах. Приведено сравнение с другими методами ранжирования вершин в графах.

В **третьей главе** представлены результаты компьютерного моделирования на примере прикладных задач анализа коммуникационных сетей. Предложенный нами способ определения электрической центральности для анализа значимости вершин рассматриваются на примере графа публикаций Math-Net.ru написанных в соавторстве, а также на ряде графов транспортных сетей Китая, Финляндии и транссибирской магистрали. Кроме того, представлен пример ранжирования сайтов научных учреждений Российской академии наук. Представлен пример применения электрической меры близости в онтологической модели компетенций.

В **заключении** представлены результаты, полученные в ходе исследования в рамках диссертационной работы.

Список использованной литературы включает 79 наименований.

Приложение А содержит листинги программного комплекса в котором проведены численные эксперименты и моделирование. Программный комплекс разработан в пакете Mathematica, а также на языке программирования Java.

Приложение Б содержит копию свидетельства о государственной регистрации программы для ЭВМ № 2015617367 от 08.07.2015.

Глава 1

Электрическая центральность вершин во взвешенных графах социальных сетей

1.1. Классическая центральность

В теории графов и анализе сетей меры центральности помогают выявлять значимость вершин внутри графа. К области применения мер центральности можно отнести, прежде всего, определение наиболее важных личностей в социальных сетях, ключевых вершин в сети Интернет и транспортных сетях, а также это важно в моделях распространения инфекционных болезней. Первоначально мера центральности использовалась в анализе социальных сетей и во многом отражали в себе идеи социологии [66].

Самой первой и самой простой мерой значимости вершин является степень вершины, т.е. число инцидентных вершине ребер или число связей, которое есть у вершины. Данная мера успешно применяется в описании процессов распространения инфекционных болезней.

Другой более сложной мерой значимости является мера центральности (betweenness centrality), первое описание было в [42]. Центральность вершины равна числу геодезических путей между всеми вершинами, проходящих через данную вершину. Центральность вершины — это важная мера, отражающая то, насколько вершина участвует в процессе распространения информации между остальными вершинами в графе.

Представим формальное определение меры центральности:

$$c_B(v) = \frac{1}{n_B} \sum_{s,t \in V} \frac{\sigma_{s,t}(v)}{\sigma_{s,t}}, \quad (1.1)$$

где $\sigma_{s,t}$ - число геодезических путей между вершинами s и t , $\sigma_{s,t}(v)$ - число геодезических путей между вершинами s и t проходящих через вершину v . Коэффициент нормировки n_B равен $n_B = (n-1)(n-2)$, если вершина v не

может быть начальной s или конечной t вершинами, и $n_B = n(n - 1)$ иначе.

Вычисление меры центральности предполагает вычисление всех геодезических путей между всеми парами вершин на графе, которая требует времени равное $O(n^3)$ с помощью алгоритма Флойда—Уоршелла. Для разреженных графов алгоритм Джонсона может быть более эффективен $O(n^2 \log n + nm)$, где m - число ребер в графе. Для невзвешенных графов наилучшая вычислительная сложность алгоритма поиска центральности равна $O(mn)$ и представлена в [28].

Пример 1.1.

Представим простой пример расчета меры центральности. На рисунке 1.1 представлен пример невзвешенного графа с пятью вершинами и пятью ребрами.

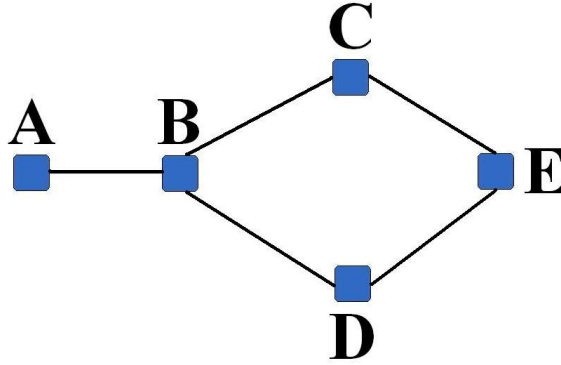


Рисунок 1.1. Пример невзвешенного графа с пятью вершинами и пятью ребрами.

Представим пример расчета центральности вершины B :

$$c_B(B) = ((\sigma_{AC}(B)/\sigma_{AC}) + (\sigma_{AD}(B)/\sigma_{AD}) + (\sigma_{AE}(B)/\sigma_{AE}) + (\sigma_{CD}(B)/\sigma_{CD}) + (\sigma_{CE}(B)/\sigma_{CE}) + (\sigma_{DE}(B)/\sigma_{DE}))/6 = ((1/1) + (1/1) + (2/2) + (1/2) + 0 + 0)/6 = 3.5/6 \approx 0.583.$$

Аналогичным образом, получаем следующие значения центральностей вершин: $c_B(A) = 0$, $c_B(B) = 7/12$, $c_B(C) = c_B(D) = 1/6$ и $c_B(E) = 1/12$.

Пример 1.2.

На рисунке 1.2 представлена карта железнодорожного сообщения Финляндии. Протяженность железных дорог Финляндии составляет 5794 км. Сеть железных дорог плотно охватывает всю страну. Пригороды столицы обслуживаются поездами ближнего радиуса¹.

¹<http://www.lvm.fi/etusivu>

Железнодорожная сеть Финляндии связывает друг с другом практически все города страны. По важнейшим маршрутам курсируют электропоезда: Хельсинки — Турку, Хельсинки — Пори, Хельсинки — Иисалми, Хельсинки — Оулу, Хельсинки — Йоенсуу и Турку — Тойяла.

Парк подвижного состава состоит из 502 дизельных локомотивов и 130 электровозов. Большинство электровозов и локомотивов оборудованы системой оповещения на основе спутниковых технологий определения местонахождения. Вагонный парк насчитывает более 14 тыс. вагонов, в том числе около 1 тыс. пассажирских вагонов².



Рисунок 1.2. Карта железнодорожного сообщения Финляндии.

Неориентированный граф, созданный на основе железнодорожного сообщения Финляндии, содержит 55 вершин и 66 ребер. Диаметр графа составляет

²Транспорт в Финляндии. (2014, декабрь 20). Википедия, свободная энциклопедия. Retrieved 09:35, августа 27, 2015 from <http://ru.wikipedia.org/?oldid=67430856>

16, средняя длина пути 5.95, число геодезических путей в графе 2970, средняя степень 2.4.

В графе учтены ребра между городами:

- Нурмес, Контиоьяки, Каджаани и Вуокатти;
- Савонлинна и Пиексямяки;
- Раума и Кокемяки;
- Пиетарсаари и Пяннинен;
- Раахе и Виханти.

Из таблицы 1.1 видно, что в первую десятку по значению центральности и степени попало 5 городов: Пиексямяки, Коуволла, Тампере, Юливиеска и Сеиняйоки. При этом первое и второе место заняли Пиексямяки и Коуволла соответственно.

Таблица 1.1

Меры центральности для вершин взвешенного графа железнодорожного сообщения Финляндии.

Вершина	Центральность c_B	Вершина	Степень
Pieksämäki	0.70417	Pieksämäki	10
Kouvola	0.527487	Kouvola	10
Tampere	0.413347	Seinäjoki	8
Iisalmi	0.401002	Oulu	7
Lahti	0.399255	Tampere	7
Mikkeli	0.396226	Vihanti	6
Ylivieska	0.375379	Ylivieska	6
Kuopio	0.373282	Kajaani	6
Seinäjoki	0.364081	Kontiomäki	6
Toijala	0.353832	Vuokatti	6
Jyväskylä	0.294782	Iisalmi	6

Oulu	0.290473	Pännäinen	6
Pännäinen	0.284184	Joensuu	6
Vihanti	0.259958	Jyväskylä	6
Tikkurila	0.247147	Parikkala	6
Parkano	0.230142	Toijala	6
Kokkola	0.225483	Lahti	6
Kemi	0.208246	Riihimäki	6
Loimaa	0.19823	Turku	6
Karjaa	0.193454	Tikkurila	6
Helsinki	0.179362	Karjaa	6
Kajaani	0.177032	Kemi	5
Turku	0.166667	Haapa-mäki	5
Haapa-mäki	0.158397	Orivesi	5
Riihimäki	0.151875	Kokemäki	5

1.2. PageRank

Кратко представим описание популярного метода ссылочного ранжирования PageRank. Идея PageRank представлена в [34]. Известно, что Google ранжирует веб-страницы в соответствии с их значениями PageRank [35].

Существует множество разных интерпретаций PageRank, например в [22] PageRank представляется как среднее число блуждающих пользователей (процессов) на данной веб-странице в данный момент времени при условии, что в каждый момент времени $t \geq 0$ каждый из пользователей может остановить свое блуждание с вероятностью $1 - \alpha$, а также в среднем $1 - \alpha$ новых пользователей начинают блуждать с любой веб-страницы. Данная интерпретация способствует более глубокому пониманию метода PageRank, но одновременно не является удобной с точки зрения практического применения, так как использует понятие момента времени.

Одной из наиболее распространенных интерпретаций можно считать интерпретацию, связанную с процессом случайного блуждания. Процесс блуждания следует по гиперссылкам с вероятностью α и с вероятностью $1 - \alpha$ может перейти на случайную веб-страницу, тогда π_i может быть интерпретирована как финальная вероятность того, что процесс случайного блуждания находится на странице i .

Математическую модель метода PageRank с точки зрения стационарного распределения из теории марковских цепей можно представить в следующем виде. Обозначим общее число вершин как n . Пусть матрица P размера $n \times n$ представлена следующим образом. Предположим, что вершина v_i имеет $k > 0$ исходящих ссылок. Тогда $p_{ij} = 1/k$ если v_j одна из исходящих ссылок и $p_{ij} = 0$ иначе. Если веб-страница не имеет исходящих ссылок, тогда $p_{ij} = 1/n$. Таким образом, PageRank определяется как стационарное распределение цепи Маркова, где пространство состояний описано множеством всех веб-страниц, и матрицей переходов:

$$\tilde{P} = \alpha P + (1 - \alpha)(1/n)E,$$

где $\alpha \in (0, 1)$, E - матрица, все элементы которой равны 1. В поисковой машине Google параметр α принимается равным 0.85. Матрица \tilde{P} является стохастической, неперiodической и неприводимой, тогда согласно эргодической теореме Маркова существует единственный вектор π такой, что $\pi \tilde{P} = \pi$, $\pi \underline{1} = 1$, где $\underline{1}$ - вектор-столбец, элементы которого равны единице.

В случае, когда используется взвешенный граф с матрицей весов W , матрица переходов P принимается равной $P = D^{-1}W$, где D - диагональная матрица степеней. Таким образом, матрицу переходов, используемую для ранжирования вершин взвешенного графа, можно представить как:

$$\tilde{P} = \alpha D^{-1}W + (1 - \alpha)(1/n)E.$$

Как мы видим, интерпретация, когда метод PageRank рассматривается, как поглощающая марковская цепь является простой и естественно описыва-

ет модель алгоритма вычисления PageRank [15]. В итоге процесс случайного блуждания начинается в некоторой случайной веб-странице и на каждом шаге может прекратиться с вероятностью $1 - \alpha$, возникновение процесса может быть описано распределением π [33, 41, 53]. Таким образом, после многократного повторения данного процесса, приближение π_j при $j = 1, \dots, n$ определяется как число случаев, когда процесс случайного блуждания заканчивается на странице j , поделенное на общее число блужданий.

Согласно доступной информации для расчетов значений PageRank компания Google использует простой метод итераций. Принимая начальное приближение равным равномерному распределению, т.е. вектор $\pi^{(0)} = (1/n)\underline{1}^T$, тогда k -ое приближение вектора можно вычислить следующим образом:

$$\pi^{(k)} = \pi^{(k-1)} \tilde{P}, k \geq 1.$$

Итерационный метод завершается после достижения некоторой заданной точности ϵ . Приблизительное число итераций необходимое для получения результата равно $\frac{\log \epsilon}{\log \alpha} nnz(P)$, где $nnz(P)$ Это число ненулевых элементов в матрице P [58]. Ряд работ предлагают способы ускорения алгоритма итераций для вычисления значения PageRank [48, 54, 55, 59].

Из других методов приближенного вычисления PageRank можно указать метод Монте-Карло. Применение метода Монте-Карло имеет ряд преимуществ:

- веб-страницы имеющие наибольшее значение PageRank определяются с высокой точностью уже после первой итерации;
- имеется естественная возможность реализации параллельных вычислений;
- метод позволяет обновлять текущие результаты расчетов PageRank с учетом изменений в структуре Всемирной сети.

Метод Монте-Карло для расчета значений PageRank представлен в следующих работах [17, 32, 41].

Отметим, что алгоритм PageRank может применяться не только к ранжированию веб-страниц, но и к любому графу.

Пример 1.3.

На рисунке 1.3 представлен пример, отражающий базовый принцип работы метода PageRank.³ На данном рисунке размеры каждой вершины пропорциональны значениям, полученным методом PageRank.

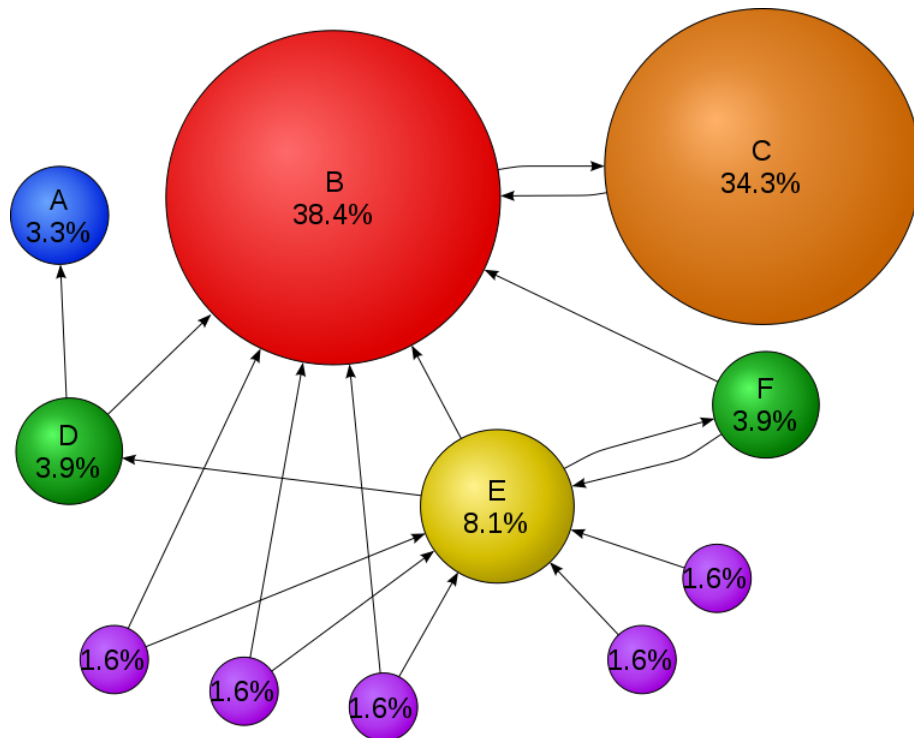


Рисунок 1.3. Базовый принцип работы PageRank.

В таблице 1.2 представлены результаты вычисления PageRank с параметром $\alpha = 0.85$ для примера из рисунка 1.3. Здесь под вершинами G, H, I, J и K промаркированы оранжевые вершины, они имеют одинаковое значение PageRank. Примечательно, что вершины D и F получили равное значение PageRank, хотя их позиции в ориентированном графе отличаются. Отметим, также что вершина E получила значение PageRank меньшее, чем вершина C , хотя имеет большое число входящих ссылок. Это объясняется тем, что вершины, ссылающиеся на вершину E , имеют малое значение PageRank, к тому же у нее больше исходящих ссылок. Совсем другая ситуация у вершины C , на

³Wikipedia contributors. "PageRank."Wikipedia, The Free Encyclopedia. Wikipedia, The Free Encyclopedia, Web. 27 Aug. 2015.

нее есть единственная ссылка, но эта ссылка является единственной исходящей ссылкой из вершины B , имеющей наибольшее значение PageRank.

Таблица 1.2

Значения PageRank с параметром $\alpha = 0.85$.

Вершина	PageRank
B	0.384401
C	0.34291
E	0.0808857
D	0.0390871
F	0.0390871
A	0.0327815
G	0.0161695
H	0.0161695
I	0.0161695
J	0.0161695
K	0.0161695

Пример 1.4.

На следующем примере представим примеры графов Европы и Азии. Здесь вершинами являются страны, а ребро это наличие сухопутных границ между соответствующими странами. Составление подобных графов применяется на уроках географии, для наглядного изучения преимуществ географического положения тех или иных стран относительно других. В граф включены страны, относящиеся к главной компоненте. По этой причине в граф стран Европы не попали такие государства как: Исландия, Ирландия, Великобритания и Мальта. Для графа стран Азии можно указать такие страны как: Япония, Тайвань, Кипр, Бахрейн, Филиппины, Сингапур, Шри-Ланка и Мальдивы. Отметим, что Россия и Турция включены как в граф стран Европы, так и в граф стран Азии.

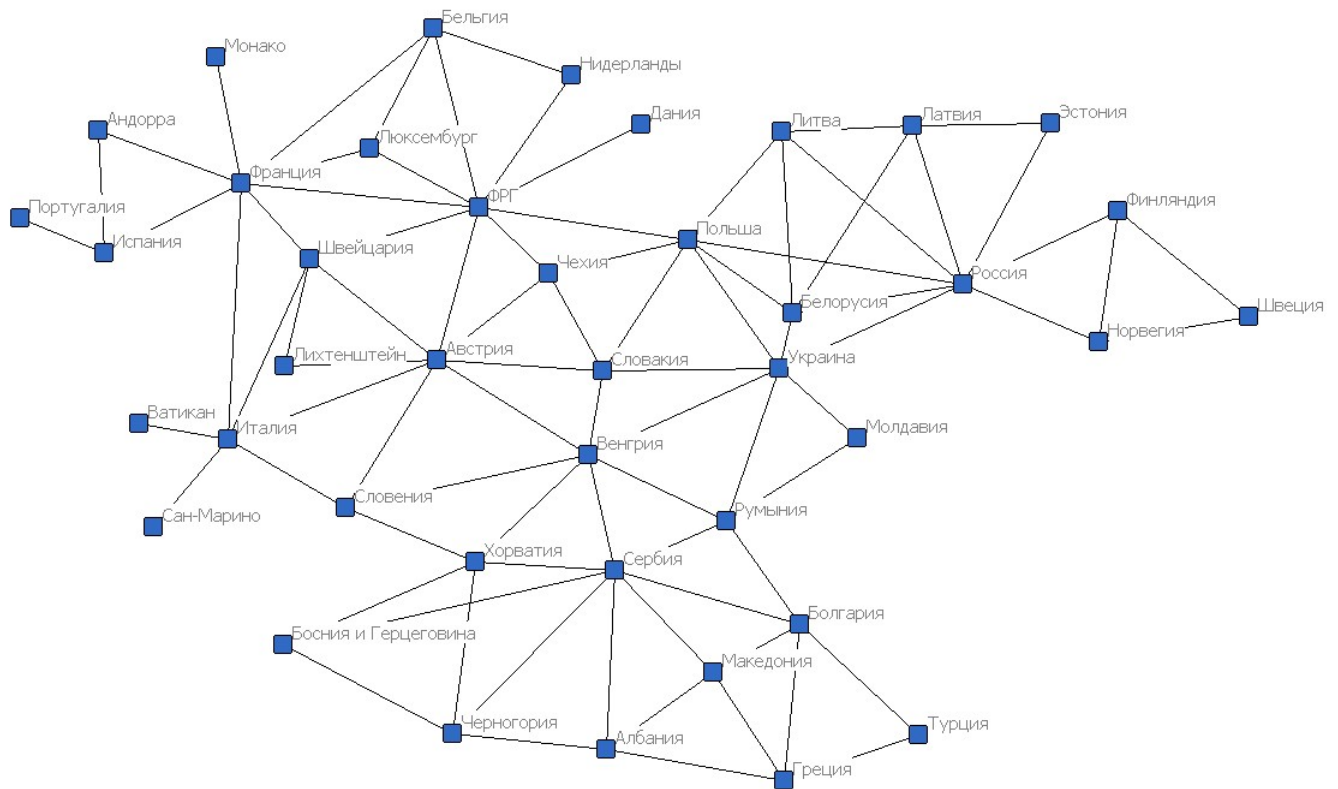


Рисунок 1.4. Граф стран Европы.

Граф стран Европы содержит 41 вершину и 84 ребра. диаметр графа: 8, средняя степень: 4.098, средняя длина пути 3.38, число геодезических путей — 1640. Результаты расчёта PageRank для графа стран Европы приведены в таблице 1.3.

Таблица 1.3

Результаты расчетов PageRank с параметром $\alpha = 0.85$ для графа стран Европы.

Вершина	PageRank $\alpha = 0.85$	Вершина	PageRank $\alpha = 0.85$
Франция	0.0522775	Словения	0.0226545
ФРГ	0.0521711	Черногория	0.022304
Россия	0.0439401	Литва	0.0221515
Австрия	0.0430856	Чехия	0.0219237
Сербия	0.0411362	Норвегия	0.0207679
Италия	0.0394101	Финляндия	0.0207679

Польша	0.035902	Люксембург	0.0195926
Украина	0.0357986	Босния и Герцеговина	0.0173204
Венгрия	0.0352756	Андорра	0.0164159
Швейцария	0.0299668	Швеция	0.015427
Болгария	0.028226	Нидерланды	0.0140381
Румыния	0.0267965	Турция	0.0135243
Хорватия	0.0267739	Лихтенштейн	0.0133307
Белорусия	0.0266715	Эстония	0.0132578
Словакия	0.0258851	Молдавия	0.0125609
Бельгия	0.0256577	Португалия	0.0108614
Испания	0.025422	Ватикан	0.00924163

Граф стран Азии содержит 40 вершин и 81 ребро. Диаметр графа: 9, средняя степень: 4.05, средняя длина пути 3.6, число геодезических путей — 1560. На рисунке 1.4 представлен граф стран Европы, а на рисунке 1.5 — граф стран Азии

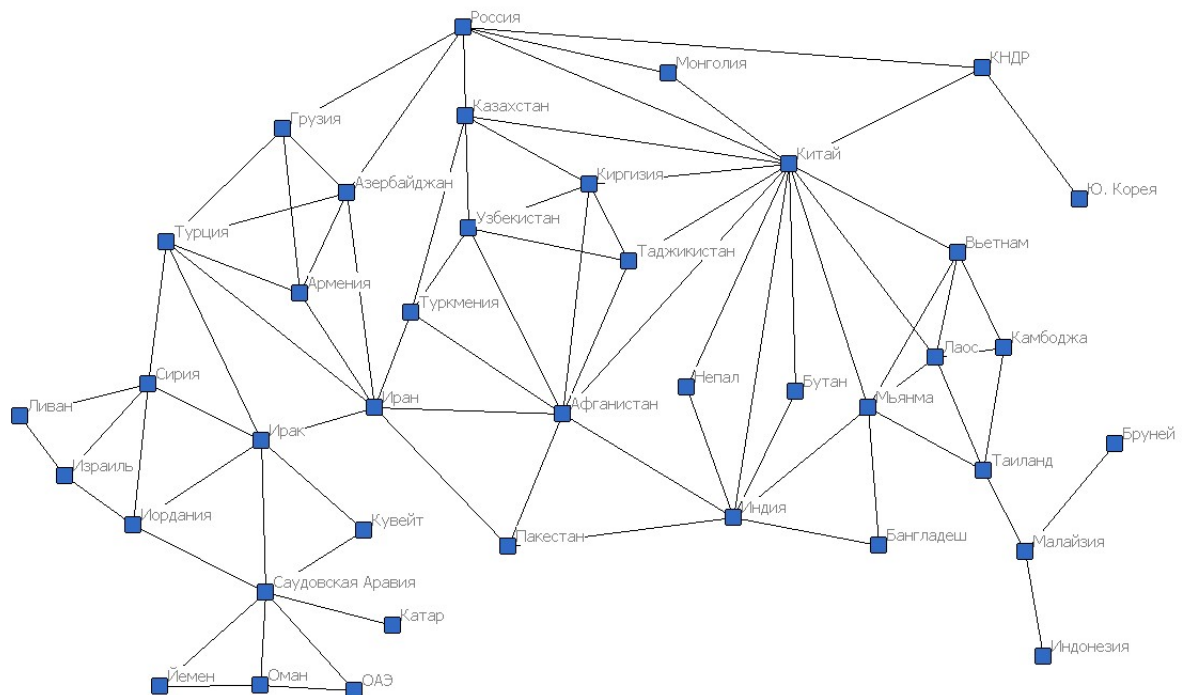


Рисунок 1.5. Граф стран Азии.

Результаты расчёта PageRank для графа стран Европы приведены в таблице 1.4.

Таблица 1.4

Результаты расчетов PageRank с параметром $\alpha = 0.85$ для графа стран Азии.

Вершина	PageRank $\alpha = 0.85$	Вершина	PageRank $\alpha = 0.85$
Китай	0.0602539	Грузия	0.0237207
Саудовская Аравия	0.0502736	Армения	0.0229079
Россия	0.0388791	Вьетнам	0.0227864
Иран	0.0372426	Таджикистан	0.0223586
Ирак	0.0365405	Туркмения	0.0221502
Индия	0.0362805	Израиль	0.0212083
Афганистан	0.0356421	Камбоджа	0.0211687
Турция	0.0335918	Пакистан	0.0170058
Сирия	0.0317544	Йемен	0.0166588
Малайзия	0.030843	ОАЭ	0.0166588
Лаос	0.0300039	Ливан	0.0151573
Азербайджан	0.0284476	Кувейт	0.0150312
Казахстан	0.0282454	Ю. Корея	0.014652
Мьянма	0.0282231	Монголия	0.0131976
Узбекистан	0.0279305	Бруней	0.0124888
Таиланд	0.0275856	Индонезия	0.0124888
Иордания	0.0264385	Бангладеш	0.0121538
Киргизия	0.0263188	Непал	0.0120952
КНДР	0.0256517	Бутан	0.0120952
Оман	0.0240146	Катар	0.00985465

Приведем также результаты вычисления значений PageRank для объединенного графа стран Евразии. Граф Евразии содержит 79 вершин и 165 ребер. Диаметр графа: 10, средняя степень: 4.177, средняя длина пути 4.315, число геодезических путей — 6162.

В таблице 1.5 представлены первые 20 стран Евразии получивших наибольшее значение PageRank. Отметим, что в первую десятку вошло больше стран относящихся только к Европе.

Таблица 1.5

Результаты расчетов PageRank с параметром $\alpha = 0.85$ для графа стран Евразии.

Вершина	PageRank $\alpha = 0.85$	Вершина	PageRank $\alpha = 0.85$
Россия	0.0393776	Украина	0.0183545
Китай	0.0294288	Иран	0.0181353
Франция	0.0270955	Ирак	0.0180564
ФРГ	0.0270115	Индия	0.0180473
Саудовская Аравия	0.0252302	Венгрия	0.0180451
Австрия	0.0222553	Афганистан	0.0176136
Турция	0.0209347	Сирия	0.015699
Сербия	0.0206099	Малайзия	0.0155761
Италия	0.0203997	Швейцария	0.0155163
Польша	0.0185025	Лаос	0.0150004

Для сравнения приведем результаты исследования, проведенные всемирным банком⁴. Всемирный банк ежегодно выводит рейтинги стран по размеру валового внутреннего продукта. Согласно рейтингу стран, уточненному в конце 2014 года⁵, в первую десятку входят: Китай, ФРГ, Франция, Италия, Россия и Турция.

⁴<http://www.worldbank.org/>

⁵Без учета стран, не вошедших в граф стран Евразии. Например, исключены из рейтинга такие страны как: США, Япония, Великобритания, Бразилия, Канада, Австралия и Мексика.

1.3. Центральность на основе правил Кирхгофа

Одним из основных недостатков классической меры центральности является то, что в расчетах учитываются только геодезические пути, и не берутся в расчет пути, которые могут быть длиннее геодезических всего на один или два шага, хотя данные негеодезические пути могут играть важную роль в описании и объяснении процессов протекающих в коммуникационных сетях.

У. Брандесом и Д. Флейсчером [29], а также М. Ньюманом [65] были предложены способы расчета центральностей, учитывающие негеодезические пути. Данные модели получили название электрическая центральность (current flow betweenness centrality) или другими словами, мера центральности на основе модели электрических цепей.

В [29, 65] граф рассматривается как электрическая цепь с идеальными элементами, где каждое ребро имеет некую пропускную способность (значение обратное сопротивлению), а вершины графа являются ее узлами. Для поиска меры центральности в модели электрической цепи используются правила Кирхгофа. Для этого электрическая цепь заземляется в некоторой вершине t и подается электрический ток в некоторой вершине s . В [29] ток подается в некоторой единственной вершине s , а также в единственной вершине t сеть заземляется. Мерой центральности вершины v служит средняя величина тока, проходящего через вершину v по всем возможным парам s и t . Таким образом, в модели электрической цепи при расчете меры центральности учитываются не только геодезические пути.

Наилучшая вычислительная сложность известного на текущий момент алгоритма поиска меры центральности в модели электрической цепи [29, 65] равна $O(I(n-1) + mn \lg n)$, где $I(n-1)$ - сложность вычисления обратной матрицы размерности $n-1$.

Далее рассмотрим новый способ определения центральности для взвешенных графов на основе модели электрических цепей.

Модель электрической цепи. В данной модели электрической цепи каждая вершина цепи соединена с искусственно вводимой вершиной v_{n+1} ребром с пропускной способностью δ . Электрический ток подается в некоторую вершину s , цепь каждый раз заземляется в одной и той же вершине v_{n+1} . Таким образом, электрический ток течет по цепи с вершины s в вершину v_{n+1} . Мерой центральности вершины v служит средняя величина тока, проходящего через вершину v по всем возможным s . Отметим, что ток на ребрах инцидентных с v_{n+1} вершиной не учитывается при расчете центральности.

Итак, представим математическое описание модели электрической цепи.

Пусть дан взвешенный граф $G = (N, E, W)$, где N множество вершин, E множество ребер, W матрица весов:

$$W(G) = \begin{pmatrix} 0 & w_{1,2} & \dots & w_{1,n} \\ w_{2,1} & 0 & \dots & w_{2,n} \\ \vdots & \vdots & \ddots & \vdots \\ w_{n,1} & w_{n,2} & \dots & 0 \end{pmatrix},$$

где $w_{i,j} \geq 0$ это вес ребра между вершинами i и j , $n = |N|$ - число вершин. Если вершины i и j не смежные, то $w_{i,j} = 0$. Пусть G - неориентированный граф, тогда $w_{i,j} = w_{j,i}$.

Далее обозначим через $D(G)$ диагональную матрицу степеней:

$$D(G) = \begin{pmatrix} d_{v_1} & 0 & \dots & 0 \\ 0 & d_{v_2} & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & d_{v_n} \end{pmatrix},$$

где $d_i = \sum_{j=1}^n w_{i,j}$ - сумма весов ребер инцидентных вершине i в графе G .

Определение 1.1. Матрица $L(G)$ называется *матрицей Кирхгофа* взвешенно-

го графа G , если

$$L(G) = D(G) - W(G) = \begin{pmatrix} d_{v_1} & -w_{1,2} & \dots & -w_{1,n} \\ -w_{2,1} & d_{v_2} & \dots & -w_{2,n} \\ \vdots & \vdots & \ddots & \vdots \\ -w_{n,1} & -w_{n,2} & \dots & d_{v_n} \end{pmatrix} \quad (1.2)$$

Пусть граф G' получен из графа G путем добавления дополнительной $n+1$ вершины соединенной со всеми вершинами графа G ребрами с постоянной проводимостью δ . Таким образом, получим следующую матрицу Кирхгофа графа G' :

$$L(G') = D(G') - W(G') = \begin{pmatrix} d_{v_1} + \delta & -w_{1,2} & \dots & -w_{1,n} & -\delta \\ -w_{2,1} & d_{v_2} + \delta & \dots & -w_{2,n} & -\delta \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ -w_{n,1} & -w_{n,2} & \dots & d_{v_n} + \delta & -\delta \\ -\delta & -\delta & \dots & -\delta & \delta n \end{pmatrix} \quad (1.3)$$

Предположим, что единица электрического тока подается в некоторую вершину $s \in N$, и вершина $n+1$ заземлена. Пусть φ_i^s обозначает абсолютный потенциал в вершине i при условии, что источник электрического тока находится в вершине s . Тогда вектор абсолютных потенциалов $\varphi^s(G') = [\varphi_1^s, \dots, \varphi_n^s, \varphi_{n+1}^s]^T$ в вершинах графа G' определяются следующей системой уравнений (правила Кирхгофа):

$$\varphi^s(G') = L(G')^{-1}b'_s, \quad (1.4)$$

где b'_s - вектор-столбец, состоящий из $n+1$ элементов, значения которых равны:

$$b'_s(i) = \begin{cases} 1 & i = s, \\ 0 & \text{иначе.} \end{cases} \quad (1.5)$$

Матрица Кирхгофа (3.2) является вырожденной. Примем в вершине $n+1$ абсолютный потенциал равным 0.

Таким образом, из (3.3) получим:

$$\tilde{\varphi}^s(G') = \tilde{L}(G')^{-1}b_s, \quad (1.6)$$

где $\tilde{\varphi}^s(G')$, $\tilde{L}(G')$ и b_s получены из (3.3) путем удаления строки и столбца, соответствующих вершине $n + 1$. Отметим, что в $\varphi^s(G')$ и b'_s удалены нулевые элементы.

Отсюда получаем

$$\tilde{\varphi}^s(G') = [D(G) - W(G) + \delta I]^{-1}b_s, \quad (1.7)$$

где I - единичная матрица размерности n .

Известно, что абсолютные потенциалы могут быть определены с точностью до постоянного слагаемого. Тогда абсолютные потенциалы $\tilde{\varphi}^s(G')$ можно принять в качестве абсолютных потенциалов в вершинах графа G , т.е.

$$\tilde{\varphi}^s(G) = [L(G) + \delta I]^{-1}b_s. \quad (1.8)$$

Представим (1.8) в следующем виде.

$$\begin{aligned} \tilde{\varphi}^s(G) &= [(D(G) + \delta I) - W(G)]^{-1}b_s = \\ &= [I - (D(G) + \delta I)^{-1}D(G)D^{-1}(G)W(G)]^{-1}(D(G) + \delta I)^{-1}b_s. \end{aligned}$$

Матрицы $(D(G) + \delta I)^{-1}$ и $(D(G) + \delta I)^{-1}D(G)$ диагональные с элементами на диагонали $\frac{1}{d_i + \delta}$ и $\frac{d_i}{d_i + \delta}$, $i = 1, \dots, n$, обозначим их D_1 и D_2 , соответственно. Матрица $D^{-1}(G)W(G)$ стохастическая, обозначим ее P . Тогда

$$\tilde{\varphi}^s(G) = [I - D_2P]^{-1}D_1b_s = \sum_{k=0}^{\infty} (D_2P)^k D_1b_s. \quad (1.9)$$

Из (1.9) следует, что потенциал можно вычислить рекуррентно в виде:

$$\tilde{\varphi}_{k+1}^s(G) = D_2P\tilde{\varphi}_k^s(G) + D_1b_s, \quad \tilde{\varphi}_0^s(G) = 0. \quad (1.10)$$

Ток, протекающий через ребро $e = (i, j)$ согласно закону Ома равен $x_e^s = |\varphi_i^s - \varphi_j^s| \cdot w_{i,j}$.

Соответственно среднее значение тока протекающего через некоторую вершину i при условии, что источник тока находится в вершине s :

$$x^s(i) = \frac{1}{2}(b_s(i) + \sum_{e:v \in e} x_e^s), \quad (1.11)$$

$$\text{где } b_s(i) = \begin{cases} 1 & i = s, \\ 0 & \text{иначе.} \end{cases}$$

Итак, меру электрической центральности в вершине i для взвешенных графов $CF_B(i)$ можно определить по формуле:

$$CF_B(i) = \frac{1}{n} \sum_{s \in N} x^s(i). \quad (1.12)$$

Таким образом, предлагаемая математическая модель электрических цепей представляет собой систему линейных алгебраических уравнений.

Вычислительная сложность алгоритма поиска электрической центральности вершин для взвешенных графов предлагаемым способом имеет относительно невысокую вычислительную сложность равную сложности вычисления обратной матрицы $O(n^3)$ (см. формулу (1.8)).

Отметим, что выражение (1.12) может быть вычислено с использованием закона больших чисел. Методы приближенного вычисления используются для упрощения и облегчения вычислительной сложности. Метод Монте-Карло может быть легко и естественно использован для получения быстрой оценки электрической центральности. Например, мы можем использовать небольшую часть вершин графа $V_1 \subset V$, выбранных последовательно независимо и равновероятно в качестве исходных для дальнейшего приближенного вычисления электрической центральности всех вершин графа:

$$CF_B(i) \approx \frac{1}{|V_1|} \sum_{s \in V_1} x^s(i). \quad (1.13)$$

Таким образом, используя только часть вершин в качестве исходных можно определить вершины с наибольшим значением электрической центральности методом Монте-Карло (1.13).

1.4. Мера близости в модели электрических цепей

В анализе социальных сетей наиболее часто используются два вида мер [76], классическая центральность (betweenness centrality) и мера близости (closeness centrality). Мера близости вершины $v \in V$ графа G впервые была представлена в [21]:

$$c_C(v) = \frac{n_C}{\sum_{v \neq t} d(v, t)}, \quad (1.14)$$

где $d(v, t)$ - кратчайший путь между вершинами v и t , n - число вершин в графе G и коэффициент нормировки $n_C = n - 1$.

Мера близости вершины $s \in V$ в рамках модели электрических цепей была предложена в [29] и определена следующим образом:

$$CF_C(s) = \frac{n_C}{\sum_{t \neq s} (\varphi^{st}(s) - \varphi^{st}(t))}, \quad (1.15)$$

где φ^{st} - абсолютный потенциал в вершине при условии, что источник тока подключен к вершине s , а сток находится в вершине t . Выражение $\varphi^{st}(s) - \varphi^{st}(t)$ соответствует действующему (эффективному) сопротивлению в электрической цепи. Таким образом, действующее сопротивление может служить некой альтернативой для расстояния между вершинами s и t .

В случае, когда абсолютные потенциалы в вершинах графа φ^s вычислены с помощью (1.8), то меру близости вершины $s \in V$ можно определить как:

$$CF_C^\delta(s) = \frac{n_C}{\sum_{v' \neq s} (\varphi^s(s) - \varphi^s(v'))}. \quad (1.16)$$

Вычислительная сложность алгоритма поиска электрической близости вершин для взвешенных графов имеет ту же вычислительную сложность, что и при поиске электрической центральности, т.е. $O(n^3)$.

1.5. Центральность вершин звезды

1.5.1. Невзвешенная звезда

Рассмотрим невзвешенный граф G с n вершинами в форме звезды. Обозначим через v_1 центр звезды.

Тогда матрица Кирхгофа $\tilde{L} = L + \delta I = D(G) - W(G) + \delta I$ имеет вид

$$L + \delta I = \begin{pmatrix} n-1+\delta & -1 & \dots & -1 \\ -1 & 1+\delta & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ -1 & 0 & \dots & 1+\delta \end{pmatrix}$$

Обратная матрица:

$$\begin{aligned} \tilde{L}^{-1} &= (L + \delta I)^{-1} = \\ &= \frac{1}{\delta(1+\delta)(n+\delta)} \begin{pmatrix} (1+\delta)^2 & 1+\delta & 1+\delta & \dots & 1+\delta \\ 1+\delta & 1+\delta(n+\delta) & 1+\delta & \dots & 1 \\ 1+\delta & 1 & 1+\delta(n+\delta) & \dots & 1 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1+\delta & 1 & 1 & \dots & 1+\delta(n+\delta) \end{pmatrix} \end{aligned}$$

Для случая $s = v_1$ по формуле (1.12) найдем

$$x^s(v_1) = \frac{1}{2} \left(1 + \frac{n-1}{n+\delta} \right),$$

$$x^s(v_i) = \frac{1}{2(n+\delta)}, \quad i = 2, \dots, n.$$

Для $s \neq v_1$

$$x^s(v_1) = \frac{2n-3+\delta}{2(1+\delta)(n+\delta)},$$

$$x^s(s) = \frac{1}{2} \left(1 + \frac{n-1+\delta}{(1+\delta)(n+\delta)} \right),$$

$$x^s(v_i) = \frac{1}{2(1+\delta)(n+\delta)}, \quad i \neq 1, s.$$

Отсюда получаем центральность вершин для звезды

$$CF_B(v_1) = \frac{1}{2n} \left(1 + \frac{n-1}{n+\delta} + (n-1) \frac{2n-3+\delta}{(1+\delta)(n+\delta)} \right) = \frac{1}{2n} + \frac{(n-1)(n-1+\delta)}{n(1+\delta)(n+\delta)},$$

$$\begin{aligned} CF_B(v_i) &= \frac{1}{2n} \left(\frac{1}{n+\delta} + 1 + \frac{n-1+\delta}{(1+\delta)(n+\delta)} + (n-2) \frac{1}{(1+\delta)(n+\delta)} \right) = \\ &= \frac{1}{2n} + \frac{n-1+\delta}{n(1+\delta)(n+\delta)}, \quad i = 2, \dots, n. \end{aligned}$$

Таким образом, мы аналитически определили электрическую центральность вершин невзвешенной звезды.

1.5.2. Звезда с одним взвешенным ребром с весом равным k

Далее рассмотрим случай, когда граф G в форме звезды с n вершинами, у которой одна из ребер имеет вес равный k , а остальные ребра имеют вес равный единице. Пусть v_1 - центр звезды, v_2 - вершина, инцидентная ребру с весом k . Тогда матрица Кирхгофа имеет вид

$$\begin{aligned} \tilde{L} &= L + \delta I = D(G) - W(G) + \delta I = \\ &= \begin{pmatrix} n-2+k+\delta & -k & -1 & \dots & -1 \\ -k & k+\delta & 0 & \dots & 0 \\ -1 & 0 & 1+\delta & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ -1 & 0 & 0 & \dots & 1+\delta \end{pmatrix} \end{aligned}$$

Обратная матрица имеет вид

$$\begin{aligned} \tilde{L}^{-1} &= (L + \delta I)^{-1} = \\ &= \frac{1}{\delta(nk + (n-1)\delta + 2k\delta + \delta^2)} \times \end{aligned}$$

$$\times \begin{pmatrix} (k+\delta)(1+\delta) & k(1+\delta) & (k+\delta) & \dots & (k+\delta) \\ k(1+\delta) & (k+(n-1)\delta+k\delta+\delta^2) & \frac{k(1+\delta)}{(1+\delta)} & \dots & k \\ (k+\delta) & k & \frac{k+\delta+\delta X}{1+\delta} & \dots & \frac{k+\delta}{1+\delta} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ (k+\delta) & k & \frac{k+\delta}{1+\delta} & \dots & \frac{k+\delta+\delta X}{1+\delta} \end{pmatrix},$$

где $X = nk + (n-1)\delta + 2k\delta + \delta^2$.

При $s = v_1$ получим следующее

$$\begin{aligned} x^s(s) &= \frac{1}{2} \left(1 + \frac{k(1+\delta) + (n-2)(k+\delta)}{nk + (n-1)\delta + 2k\delta + \delta^2} \right), \\ x^s(2) &= \frac{k(1+\delta)}{2(nk + (n-1)\delta + 2k\delta + \delta^2)}, \\ x^s(i) &= \frac{k+\delta}{2(nk + (n-1)\delta + 2k\delta + \delta^2)}, \quad i = 3, \dots, n. \end{aligned}$$

При $s = v_2$

$$\begin{aligned} x^s(1) &= \frac{k(\delta + 2n - 3)}{2(nk + (n-1)\delta + 2k\delta + \delta^2)}, \\ x^s(s) &= \frac{1}{2} \left(1 + \frac{k(\delta + n - 1)}{nk + (n-1)\delta + 2k\delta + \delta^2} \right), \\ x^s(i) &= \frac{k}{2(nk + (n-1)\delta + 2k\delta + \delta^2)}, \quad i = 3, \dots, n. \end{aligned}$$

При $s = v_i$, где $i = 3, \dots, n$

$$\begin{aligned} x^s(v_1) &= \frac{1}{2} \frac{\delta^2 + \delta(3k + 2n - 5) + k(2n - 3)}{(nk + (n-1)\delta + 2k\delta + \delta^2)(1+\delta)}, \\ x^s(v_2) &= \frac{k}{2(nk + (n-1)\delta + 2k\delta + \delta^2)}, \\ x^s(s) &= \frac{1}{2} \left(1 + \frac{1}{1+\delta} - \frac{k+\delta}{(nk + (n-1)\delta + 2k\delta + \delta^2)(1+\delta)} \right), \\ x^s(v_i) &= \frac{k+\delta}{2(nk + (n-1)\delta + 2k\delta + \delta^2)(1+\delta)}, \quad i = 3, \dots, n. \end{aligned}$$

Таким образом, получаем следующий результат

$$CF_B(v_1) = \frac{1}{2n} \left(1 + \frac{2k(\delta + n - 1)}{X} + \frac{2(n - 2)(\delta^2 + \delta(2k + n - 2) + k(n - 1))}{(1 + \delta)X} \right),$$

$$CF_B(v_2) = \frac{1}{2n} \left(1 + \frac{2k(\delta + n - 1)}{X} \right),$$

$$CF_B(v_i) = \frac{1}{2n} \left(1 + \frac{1}{1 + \delta} + \frac{(k + \delta)(n - 4)}{(1 + \delta)X} + \frac{2k + \delta}{X} \right), \quad i = 3, \dots, n.$$

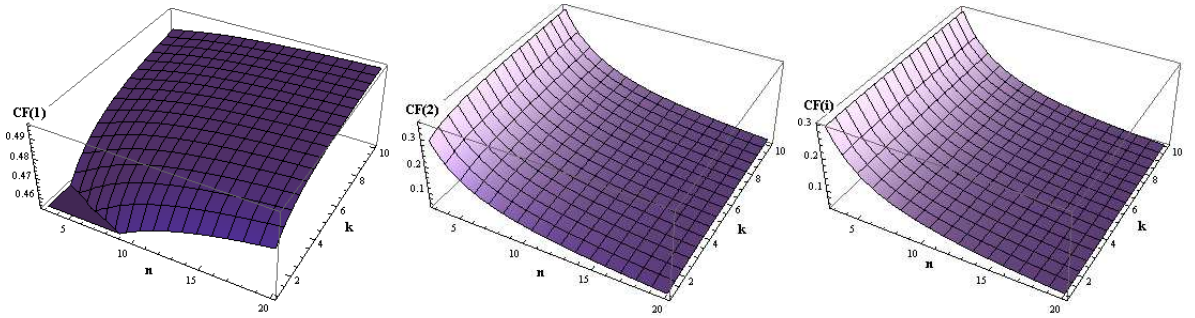


Рисунок 1.6. Зависимость значений $CF_B(1)$, $CF_B(2)$ и $CF_B(i)$ от n и k при $\delta = 1$.

На рис. 1.6 представлены графики зависимостей значений электрической центральности от значений n и k при $\delta = 1$. С увеличением значений n и k центральность $CF_B(1)$ растет. С другой стороны с увеличением n центральности $CF_B(2)$ и $CF_B(i)$ уменьшается. Изменение k незначительно влияет на изменение центральностей $CF_B(2)$ и $CF_B(i)$.

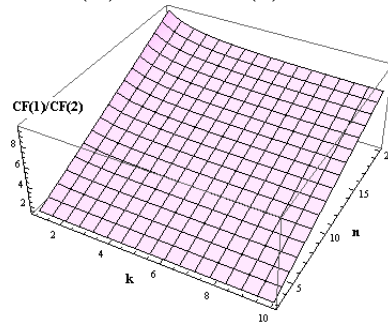


Рисунок 1.7. Зависимость соотношения центральностей $\frac{CF_B(1)}{CF_B(2)}$ от n и k при $\delta = 1$.

1.6. Центральность вершин полного двудольного графа

Рассмотрим частный случай, когда G - полный двудольный граф с n вершинами, вершины которого разделены на два множества V_1 и V_2 таким образом, что

ни одна вершина в V_1 не соединена с вершинами V_1 , и ни одна вершина в V_2 не соединена с вершинами V_2 . Обозначим полный двудольный граф как $K_{|V_1|,|V_2|}$. Все ребра имеют вес равный единице. Для случая $K_{1,n-1}$, т.е. для звезды аналитическое решение электрической центральности представлено ранее.

1.6.1. Двудольный граф $K_{2,n-2}$.

Пусть $V_1 = \{v_1, v_2\}$, $v' \in V_2$. Тогда матрица Кирхгофа имеет вид:

$$D(G) - W(G) + \delta I =$$

$$= \begin{pmatrix} n-2+\delta & 0 & -1 & -1 & \dots & -1 \\ 0 & n-2+\delta & -1 & -1 & \dots & -1 \\ -1 & -1 & 2+\delta & 0 & \dots & 0 \\ -1 & -1 & 0 & 2+\delta & \dots & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ -1 & -1 & 0 & 0 & \dots & 2+\delta \end{pmatrix}.$$

Обратная матрица имеет вид:

$$(D(G) - W(G) + \delta I)^{-1} =$$

$$= \frac{1}{\delta(n+\delta)} \begin{pmatrix} \frac{n-2+n\delta+\delta^2}{n+\delta-2} & \frac{n-2}{n+\delta-2} & 1 & 1 & \dots & 1 \\ \frac{n-2}{n+\delta-2} & \frac{n-2+n\delta+\delta^2}{n+\delta-2} & 1 & 1 & \dots & 1 \\ 1 & 1 & \frac{2+n\delta+\delta^2}{2+\delta} & \frac{2}{2+\delta} & \dots & \frac{2}{2+\delta} \\ 1 & 1 & \frac{2}{2+\delta} & \frac{2+n\delta+\delta^2}{2+\delta} & \dots & \frac{2}{2+\delta} \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & 1 & \frac{2}{2+\delta} & \frac{2}{2+\delta} & \dots & \frac{2+n\delta+\delta^2}{2+\delta} \end{pmatrix}.$$

При $s = v_1$ ток распределяется следующим образом:

$$x^s(v_1) = \frac{1}{2} \left(1 + \frac{(\delta + n - 1)(n - 2)}{(n + \delta)(n + \delta - 2)} \right),$$

$$x^s(v_2) = \frac{(n - 2)}{2(n + \delta - 2)(\delta + n)},$$

$$x^s(v') = \frac{1}{2(n + \delta - 2)}.$$

Так как вершины v_1 и v_2 расположены симметрично, то при $s = v_2$ получаем:

$$\begin{aligned} x^s(v_1) &= \frac{(n-2)}{2(n+\delta-2)(\delta+n)}, \\ x^s(v_2) &= \frac{1}{2} \left(1 + \frac{(\delta+n-1)(n-2)}{(n+\delta)(n+\delta-2)} \right), \\ x^s(v') &= \frac{1}{2(n+\delta-2)}. \end{aligned}$$

При $s = v'$:

$$\begin{aligned} x^s(v_1) &= x^s(v_2) = \frac{\delta + 2n - 4}{2(2 + \delta)(\delta + n)}, \\ x^s(s) &= \frac{1}{2} \left(1 + \frac{2(\delta + n - 1)}{(2 + \delta)(\delta + n)} \right), \\ x^s(v') &= \frac{1}{(2 + \delta)(\delta + n)}. \end{aligned}$$

Получаем следующий результат для $K_{2,n-2}$:

$$\begin{aligned} CF_B(v_1) &= CF_B(v_2) = \frac{1}{2n} \left(1 + \frac{n-2}{n+\delta-2} + \frac{(n-2)(\delta+2n-4)}{(2+\delta)(\delta+n)} \right), \\ CF_B(v') &= \frac{1}{2n} \left(1 + \frac{2}{n+\delta-2} + \frac{2(\delta+2n-4)}{(2+\delta)(\delta+n)} \right). \end{aligned}$$

1.6.2. Двудольный граф $K_{3,n-3}$.

Пусть $V_1 = \{v_1, v_2, v_3\}$, $v' \in V_2$. Тогда матрица Кирхгофа:

$$D(G) - W(G) + \delta I =$$

$$= \begin{pmatrix} n-3+\delta & 0 & 0 & -1 & -1 & \dots & -1 \\ 0 & n-3+\delta & 0 & -1 & -1 & \dots & -1 \\ 0 & 0 & n-3+\delta & -1 & -1 & \dots & -1 \\ -1 & -1 & -1 & 3+\delta & 0 & \dots & 0 \\ -1 & -1 & -1 & 0 & 3+\delta & \dots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ -1 & -1 & -1 & 0 & 0 & \dots & 3+\delta \end{pmatrix}.$$

Обратная матрица имеет вид:

$$(D(G) - W(G) + \delta I)^{-1} = \frac{1}{\delta(n+\delta)} \times$$

$$\times \begin{pmatrix} \frac{n-3+n\delta+\delta^2}{n+\delta-3} & \frac{n-3}{n+\delta-3} & \frac{n-3}{n+\delta-3} & 1 & 1 & \dots & 1 \\ \frac{n-3}{n+\delta-3} & \frac{n-3+n\delta+\delta^2}{n+\delta-3} & \frac{n-3}{n+\delta-3} & 1 & 1 & \dots & 1 \\ \frac{n-3}{n+\delta-3} & \frac{n-3}{n+\delta-3} & \frac{n-3+n\delta+\delta^2}{n+\delta-3} & 1 & 1 & \dots & 1 \\ 1 & 1 & 1 & \frac{3+n\delta+\delta^2}{3+\delta} & \frac{3}{3+\delta} & \dots & \frac{3}{3+\delta} \\ 1 & 1 & 1 & \frac{3}{3+\delta} & \frac{3+n\delta+\delta^2}{3+\delta} & \dots & \frac{3}{3+\delta} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & 1 & 1 & \frac{3}{3+\delta} & \frac{3}{3+\delta} & \dots & \frac{3+n\delta+\delta^2}{3+\delta} \end{pmatrix}.$$

При $s = v_1$ ток распределяется следующим образом:

$$x^s(s) = \frac{1}{2} \left(1 + \frac{(\delta + n - 1)(n - 3)}{(n + \delta)(n + \delta - 3)} \right),$$

$$x^s(v_2) = x^s(v_3) = \frac{n - 3}{2(n + \delta)(n + \delta - 3)},$$

$$x^s(v') = \frac{\delta + n + 1}{2(n + \delta)(n + \delta - 3)}.$$

Так как вершины v_1 , v_2 и v_3 расположены симметрично, то при $s = v_2$ получаем:

$$x^s(s) = \frac{1}{2} \left(1 + \frac{(\delta + n - 1)(n - 3)}{(n + \delta)(n + \delta - 3)} \right),$$

$$x^s(v_1) = x^s(v_3) = \frac{n-3}{2(n+\delta)(n+\delta-3)},$$

$$x^s(v') = \frac{\delta+n+1}{2(n+\delta)(n+\delta-3)}.$$

а при $s = v_3$ получаем:

$$x^s(s) = \frac{1}{2} \left(1 + \frac{(\delta+n-1)(n-3)}{(n+\delta)(n+\delta-3)} \right),$$

$$x^s(v_1) = x^s(v_2) = \frac{n-3}{2(n+\delta)(n+\delta-3)},$$

$$x^s(v') = \frac{\delta+n+1}{2(n+\delta)(n+\delta-3)}.$$

При $s = v'$:

$$x^s(v_1) = x^s(v_2) = x^s(v_3) = \frac{\delta+2n-5}{2(3+\delta)(\delta+n)},$$

$$x^s(s) = \frac{1}{2} \left(1 + \frac{3(\delta+n-1)}{(3+\delta)(\delta+n)} \right),$$

$$x^s(v') = \frac{3}{2(3+\delta)(\delta+n)}.$$

Получаем следующий результат для $K_{3,n-3}$:

$$CF_B(v_1) = CF_B(v_2) = CF_B(v_3) =$$

$$= \frac{1}{2n} \left(1 + \frac{(n-3)(\delta+n+1)}{(n+\delta-3)(\delta+n)} + \frac{(n-3)(\delta+2n-5)}{(3+\delta)(\delta+n)} \right),$$

$$CF_B(v') = \frac{1}{2n} \left(1 + \frac{3(\delta+n+1)}{(n+\delta-3)(\delta+n)} + \frac{3(\delta+2n-5)}{(3+\delta)(\delta+n)} \right).$$

1.6.3. Двудольный граф $K_{r,n-r}$.

Пусть $v \in V_1$, $v' \in V_2$ и $r = |V_1|$, $n-r = |V_2|$. Используя индукцию, из результатов для $K_{1,n-1}$, $K_{2,n-2}$ и $K_{3,n-3}$ можно определить значения центральностей для вершин полного двудольного графа $K_{r,n-r}$:

При $r = 1$, т.е. при $K_{1,n-1}$:

$$CF_B(v) = \frac{1}{2n} \left(1 + \frac{2(n-1)(n-1+\delta)}{(1+\delta)(\delta+n)} \right),$$

$$CF_B(v') = \frac{1}{2n} \left(1 + \frac{2(\delta+n-1)}{(1+\delta)(\delta+n)} \right).$$

При $r = 2$ ($K_{2,n-2}$):

$$CF_B(v) = \frac{1}{2n} \left(1 + \frac{n-2}{n+\delta-2} + \frac{(n-2)(\delta+2n-4)}{(2+\delta)(\delta+n)} \right),$$

$$CF_B(v') = \frac{1}{2n} \left(1 + \frac{2}{n+\delta-2} + \frac{2(\delta+2n-4)}{(2+\delta)(\delta+n)} \right).$$

При $r = 3$ ($K_{3,n-3}$):

$$CF_B(v) = \frac{1}{2n} \left(1 + \frac{(n-3)(\delta+n+1)}{(n+\delta-3)(\delta+n)} + \frac{(n-3)(\delta+2n-5)}{(3+\delta)(\delta+n)} \right),$$

$$CF_B(v') = \frac{1}{2n} \left(1 + \frac{3(\delta+n+1)}{(n+\delta-3)(\delta+n)} + \frac{3(\delta+2n-5)}{(3+\delta)(\delta+n)} \right).$$

Итак, $K_{r,n-r}$:

$$CF_B(v) = \frac{1}{2n} \left(1 + \frac{(n-r)(\delta+n-2+r)}{(n+\delta-r)(\delta+n)} + \frac{(n-r)(\delta+2n-2-r)}{(r+\delta)(\delta+n)} \right),$$

$$CF_B(v') = \frac{1}{2n} \left(1 + \frac{r(\delta+n-2+r)}{(n+\delta-r)(\delta+n)} + \frac{r(\delta+2n-2-r)}{(r+\delta)(\delta+n)} \right).$$

Таким образом, мы аналитически определили электрическую центральность вершин полного двудольного графа.

1.7. Сравнение с PageRank и классической центральностью

1.7.1. Простой пример взвешенного графа

В качестве простого примера рассмотрим взвешенный граф, со следующей матрицей весов:

$$\begin{array}{c} A \quad B \quad C \quad D \\ \begin{array}{l} A \\ B \\ C \\ D \end{array} \begin{pmatrix} 0 & 1 & 3 & 1 \\ 1 & 0 & 2 & 0 \\ 3 & 2 & 0 & 0 \\ 1 & 0 & 0 & 0 \end{pmatrix} \end{array}.$$

Данный взвешенный граф представлен на рисунке 1.8. При расчете классической меры центральности, т.е. без учета весов, вершины B и C получают равное значение меры центральности (см. таблица 1). Но в случае, когда вес ребра учитывается, интуитивно можно предположить, что вершина C имеет большее значение центральности по сравнению с вершиной B , так как у вершины C более сильная связь с вершиной A . А относительно связей с остальными вершинами у вершин B и C равные позиции.

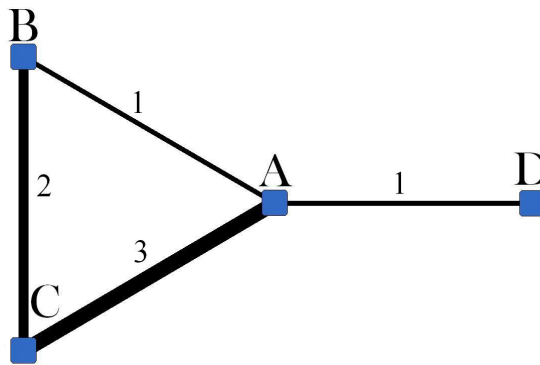


Рисунок 1.8. Простой пример взвешенного графа.

Результаты вычислений классической меры центральности, меры электрической центральности с параметром $\delta = 1$ и значений PageRank с параметром $\alpha = 0.85$ приведены в таблице 1.6. Данные результаты подтверждают сделанный нами ранее интуитивный вывод: вершина C имеет большее значение центральности, чем вершина B .

Таблица 1.6

Меры центральности для взвешенного графа на рис. 1.8.

Вершина	A	B	C	D
Классическая центральность (betweenness centrality)	2	0	0	0
Мера центральности PageRank ($\alpha = 0.85$)	0.352665	0.212248	0.337633	0.0974531
Электрическая центральность ($\delta = 1$)	0.401786	0.267857	0.334821	0.227679

1.7.2. Пример взвешенного графа с 6 вершинами

Рассмотрим пример взвешенного графа с 6 вершинами (см. рисунок 1.9) с матрицей весов:

$$\begin{matrix} & A & B & C & D & E & F \\ \begin{matrix} A \\ B \\ C \\ D \\ E \\ F \end{matrix} & \begin{pmatrix} 0 & 1 & 1 & 1 & 0 & 0 \\ 1 & 0 & 2 & 0 & 0 & 0 \\ 1 & 2 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 1 & 0 & 2 \\ 0 & 0 & 0 & 1 & 2 & 0 \end{pmatrix} \end{matrix} .$$

Результаты ранжирования вершин данного графа наглядно показывает нам сильные стороны электрической центральности.

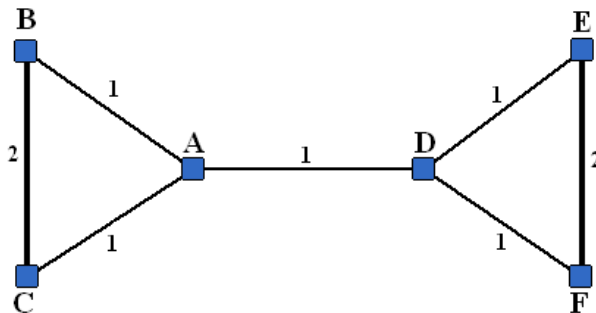


Рисунок 1.9. Взвешенный граф с 6 вершинами.

В таблице 1.7 представлены результаты вычисления мер центральностей: классическая центральность, электрическая центральность с параметром $\delta = 1$ и мера центральности PageRank с параметром $\alpha = 0.85$.

Здесь мы видим, что классическая центральность выделяет вершины A и D , но дает нулевую центральность остальным вершинам. Метод ранжирования PageRank всем вершинам дает одинаковое значение центральности. С другой стороны, электрическая центральность выделяет центральность вершин A и D относительно центральностей остальных вершин.

Таблица 1.7

Меры центральности для взвешенного графа из 6 вершин.

Вершины	A	B	C	D	E	F
Классическая центральность (betweenness centrality)	6	0	0	6	0	0
PageRank центральность $\alpha = 0.85$	1/6	1/6	1/6	1/6	1/6	1/6
Электрическая центральность $\delta = 1$	0.2708	0.191	0.191	0.2708	0.191	0.191

1.7.3. Электрическая центральность двух звезд

В данном разделе рассмотрим пример графа состоящего из двух звезд соединенных между собой одной вершиной. Граф представлен на рисунке 1.10.

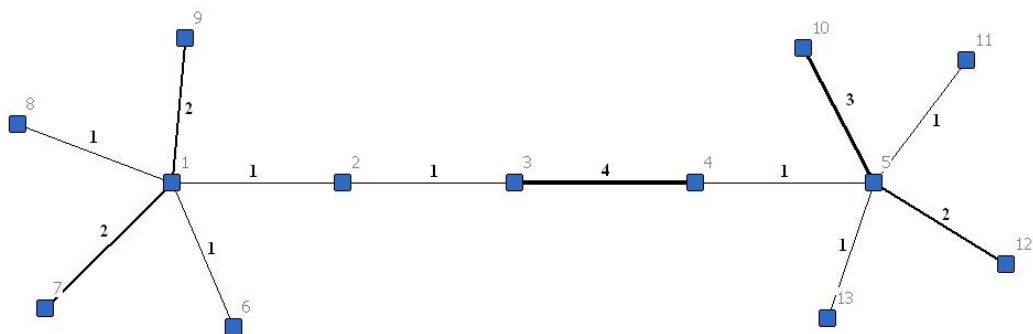


Рисунок 1.10. Две соединённые звезды.

Если рассматривать центральности вершин представленного графа без учета весов ребер ясно, что вершины 1 и 5 должны иметь максимальное значение центральности. Далее вершины 2 , 3 и 4 должны иметь значение центральности большее, чем вершины-листья: 6 , 7 , 8 , 9 , 10 , 11 , 12 и 13 . Отметим, что центральности вершин 2 и 4 должны быть равны.

Приведём результаты расчёта электрической центральности и PageRank для графа без учёта весов рёбер приведены в таблице 1.8.

Таблица 1.8

Результаты расчетов электрической центральности с параметром $\delta = 1$ и PageRank с параметром $\alpha = 0.85$.

Вершина	Центральность (CF_B)	Вершина	PageRank $\alpha = 0.85$
1	0.21049	1	0.199779
5	0.21049	5	0.199779
2	0.113147	2	0.0789114
4	0.113147	4	0.0789114
3	0.108951	3	0.0786132
6	0.0716084	6	0.0455008
7	0.0716084	7	0.0455008
8	0.0716084	8	0.0455008
9	0.0716084	9	0.0455008
10	0.0716084	10	0.0455008
11	0.0716084	11	0.04550082
12	0.0716084	12	0.0455008
13	0.0716084	13	0.0455008

Из таблицы 1.8 видно, что результаты ранжирования электрической центральности с параметром $\delta = 1$ и PageRank с параметром $\alpha = 0.85$ совпадают. Коэффициент корреляции значений электрической центральности и значений

PageRank равен 0.99602.

При учете весов ребер можно сделать следующий интуитивный вывод о центральности вершин графа. Вершины 1 и 5 также должны получить высокое значение центральности относительно остальных вершин графа, но вершина 5 имеет небольшое преимущество перед вершиной 1, т.к. у нее имеется более весомое ребро с вершиной 10.

Что касается центральности вершин 2 и 4, то вершина 4 имеет преимущество перед вершиной 2 за счет ребра с вершиной 3 с весом 4. Также ясно, что из вершин-листьев графа наибольшую центральность должна получить вершина 10, далее идут вершины 12, 7 и 9. При этом центральность вершины 12 должна быть выше центральности вершин 7 и 9 за счет близости к ребру с весом 3.

В таблице 1.9 приведены результаты вычислений мер центральности для двух соединённых звёзд на рисунке 3 с учётом весов рёбер.

Таблица 1.9

Результаты расчетов электрической центральности с параметром $\delta = 1$ и PageRank с параметром $\alpha = 0.85$.

Вершина	Центральность (CF_B)	Вершина	PageRank ($\alpha = 0.85$)
5	0.239918	5	1.25188
1	0.231646	1	0.999936
4	0.136309	4	0.794229
3	0.130677	3	0.726229
2	0.117323	10	0.529138
10	0.0856016	2	0.396896
12	0.081406	12	0.379426
7	0.0811138	7	0.340832
9	0.0811138	9	0.340832
11	0.0722332	11	0.203046
13	0.0722332	13	0.203046

6	0.0720688	6	0.183749
8	0.0720688	8	0.183749

Из таблицы 1.9 видно, что результаты ранжирования электрической центральности с параметром $\delta = 1$ и PageRank с параметром $\alpha = 0.85$ не совпадают только относительно вершин 2 и 10. В остальном же результаты ранжирования совпадают со сделанными ранее интуитивными выводами о центральности вершин графа.

1.8. Приложение для социальных сетей

В настоящем примере представлен взвешенный граф, полученный из известной социальной сети ВКонтакте. Граф получен путем исследования самого большого сообщества, посвященного теории игр состоящего из 483 участников. В качестве веса ребра в графе принято число общих друзей между участниками. Действительно, вероятность того, что некоторые два участника знакомы друг с другом зависит от числа общих друзей. Данный подход широко используется для поиска друзей практически во всех современных социальных сетях.

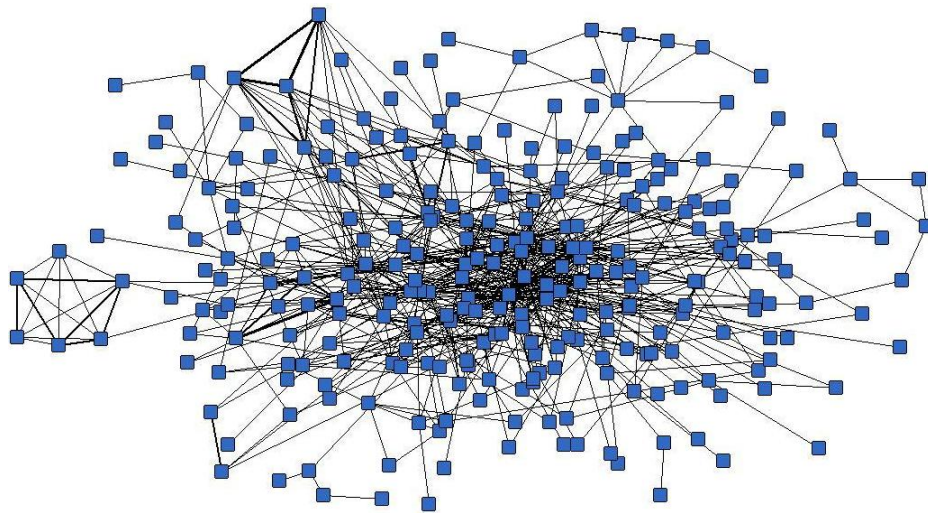


Рисунок 1.11. Главная компонента графа сообщества «Теория игр» в социальной сети ВКонтакте (275 вершин).

На рисунке 1.11 представлена главная компонента данного графа, состоящая из 275 вершин, 805 ребер, диаметр графа равен 8. Из рисунка 1.11 видно,

что данный граф является неудобным для наглядного, визуального исследования. Поэтому рассмотрим взвешенный граф, полученный из графа, представленного на рисунке 1.12, путем удаления более слабых связей. Удаление ребер с весом меньше 3 существенно упрощает граф. Главная компонента графа, получившегося после удаления ребер с весом меньше 3, представлена на рисунке 1.12, и состоит уже из 71 вершин и 116 ребер, диаметр графа увеличился до 9. Толщина ребра на рисунках 1.11 и 1.12 зависит от веса ребра, т.е. от количества общих друзей.

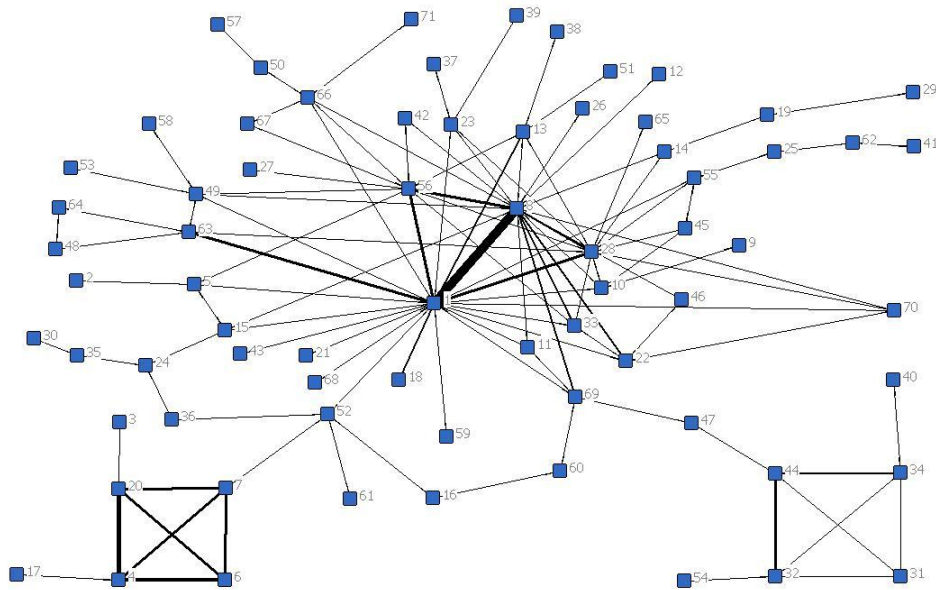


Рисунок 1.12. Главная компонента графа сообщества «Теория игр» в социальной сети ВКонтакте после удаления ребер с весом меньше 3 (71 вершина).

Из рисунка 1.12 видно, что вершина 52 соединяет подграф $\{3, 4, 6, 7, 17, 20\}$ с остальной частью графа. Отсюда можно сделать интуитивный вывод о том, что центральность вершины 52 должна быть выше вершин подграфа $\{3, 4, 6, 7, 17, 20\}$. Отметим, что вершина 7 также должна иметь более высокую центральность относительно вершин 3, 4, 6, 17 и 20.

Результаты вычислений меры центральности на основе правил Кирхгофа для взвешенного графа, представленного на рисунке 1.11, приведены в таблице 1.10. Расчеты проведены с параметром $\delta = 0.3$. В таблице 1.11 приведены результаты вычислений центральности для того же графа методом PageRank с параметром $\alpha = 0.85$.

Результаты расчетов меры электрической центральности на основе правил Кирхгофа для взвешенного графа с параметром $\delta = 0.3$.

Вершина	Электрическая центральность (CF_B)	Вершина	Электрическая центральность (CF_B)
1	0.416834	11	0.0293603
8	0.314303	67	0.027035
52	0.146293	48	0.0264221
69	0.145402	70	0.0233142
28	0.129854	64	0.0232039
56	0.127302	31	0.0215073
7	0.100162	18	0.0207139
15	0.0931205	46	0.020683
66	0.0922263	27	0.0205176
63	0.089614	12	0.0205118
47	0.0888958	42	0.020321
24	0.0880064	37	0.0202657
44	0.0841683	39	0.020204
55	0.0800937	58	0.0202029
49	0.0724918	68	0.0201055
23	0.0702061	2	0.0199689
13	0.069915	9	0.0199592
10	0.0609674	53	0.0199322
14	0.0598266	71	0.0199179
25	0.0563891	38	0.0198471
60	0.0543068	61	0.0197537
4	0.0509786	57	0.0196864

36	0.0500803	21	0.0196383
32	0.0500185	43	0.0196383
5	0.0481911	59	0.0196383
20	0.0480721	26	0.0196361
16	0.0469153	65	0.019613
34	0.0465511	51	0.0195626
22	0.044108	30	0.0192754
50	0.0430987	17	0.0190127
45	0.0417735	3	0.0190104
35	0.0409987	40	0.0189889
62	0.0398752	41	0.0189866
19	0.0392692	54	0.0187614
6	0.0344222	29	0.0184664
33	0.0313602		

Таблица 1.11

Результаты расчетов меры центральности PageRank с параметром $\alpha = 0.85$.

Вершина	PageRank ($\alpha = 0.85$)	Вершина	PageRank ($\alpha = 0.85$)	Вершина	PageRank ($\alpha = 0.85$)
1	0.135924	50	0.0122715	27	0.00553029
8	0.118936	55	0.011047	58	0.00550747
56	0.0432228	30	0.010792	67	0.00518201
28	0.0366455	15	0.010617	12	0.00507816
44	0.0277399	5	0.0100218	61	0.00483279
4	0.0267178	31	0.00978122	47	0.00453329
32	0.0251955	25	0.0097015	71	0.00438022
20	0.0244063	57	0.00928381	53	0.00423442
63	0.0227781	41	0.00925256	46	0.00403814
6	0.021239	45	0.00922356	9	0.00401267

34	0.0204177	19	0.00910172	42	0.00377585
7	0.0204087	36	0.00902109	40	0.00362181
22	0.019378	18	0.00898979	68	0.0034881
13	0.0185616	70	0.00822758	38	0.00348462
23	0.0170018	14	0.00821881	54	0.00318349
33	0.0165097	11	0.00776778	51	0.00314164
49	0.0159755	37	0.00721322	65	0.00300264
62	0.0157497	16	0.00674012	21	0.00293793
69	0.0149976	48	0.00642346	43	0.00293793
66	0.0144056	39	0.00636313	59	0.00293793
35	0.0139241	60	0.00598379	26	0.00292144
52	0.0134405	29	0.00598091	3	0.00252758
10	0.0129647	2	0.0056203	17	0.0025231
24	0.0124149	64	0.00558741		

Результаты ранжирования электрической центральностью (CF_B) подтверждают сделанные нами интуитивные выводы. Вершины 52 и 7 получили 3 и 7 места соответственно. Вершины 4, 20, 6, 17 и 3 заняли соответственно 22, 24, 36, 68 и 69 места.

Результаты ранжирования PageRank дают результаты отличные от наших интуитивных выводов. Вершины 52 и 7 получили 22 и 12 места, а вершины 4, 20, 6 заняли соответственно 6, 8, 10 места. Т.е. получили значения центральности выше, чем вершины 52 и 7.

Применим стохастический метод Монте-Карло для приближенного вычисления электрической центральности с помощью выражения (1.13).

В таблице 1.12 представлены результаты приближенного вычисления электрической центральности вершин графа с помощью метода Монте-Карло, в котором использовались в качестве исходных только 10% вершин. Вершины V_1 выбраны последовательно независимо и равновероятно.

Таблица 1.12

Результаты применения метода Монте-Карло с использованием 10% вершин для расчета электрической центральности с параметром $\delta = 0.3$.

Вершина	Элект. центр-сть (CF_B) $\delta = 0.3$	Вершина	Метод Монте-Карло 10% вершин
1	0.4168	1	0.5043
8	0.3143	8	0.4134
52	0.1463	52	0.2468
69	0.1454	23	0.2307
28	0.1299	28	0.2255
56	0.1273	20	0.2003
7	0.1002	7	0.1982
15	0.0931	24	0.1871
66	0.0922	63	0.1789
63	0.0896	10	0.1786
47	0.0889	15	0.1763
24	0.0880	55	0.1756
44	0.0842	36	0.1613
55	0.0801	69	0.1565
49	0.0725	12	0.1457
23	0.0702	39	0.1438
13	0.0699	45	0.1403
10	0.0610	56	0.1397
14	0.0598	3	0.1360
25	0.0564	4	0.1234

Как и предполагалось ранее, вершины с наибольшим значением центральности 1 и 8 выявлены. Помимо этого 6 вершин из первой десятки также бы-

ли выявлены. Метод Монте-Карло позволил также правильно определить ранг вершин 52 и 7.

Таблица 1.13

Результаты применения метода последовательных приближений для расчета электрической центральности с параметром $\delta = 0.3$.

Вершина	Электр. центр-сть (CF_B)	Вершина	Метод итераций ($k = 2$)	Вершина	Метод итераций ($k = 100$)
1	0.416834	1	0.178304	1	0.371244
8	0.314303	8	0.145592	8	0.284146
52	0.146293	56	0.0659887	52	0.127214
69	0.145402	28	0.0572842	69	0.123302
28	0.129854	13	0.0431609	28	0.118431
56	0.127302	63	0.0415547	56	0.11654
7	0.100162	49	0.0368692	7	0.0953179
15	0.0931205	66	0.0347638	15	0.0819878
66	0.0922263	52	0.0344588	63	0.0811178
63	0.089614	34	0.0342016	66	0.0808894
47	0.0888958	4	0.0338743	24	0.0766251
24	0.0880064	44	0.0335074	47	0.0740926
44	0.0841683	32	0.0333317	44	0.0730975
55	0.0800937	20	0.0327954	55	0.0704496
49	0.0724918	69	0.0326849	49	0.0636043
23	0.0702061	22	0.0314009	23	0.0622796
13	0.069915	10	0.0305125	13	0.0613457
10	0.0609674	23	0.0303764	10	0.0543806
14	0.0598266	7	0.0260482	14	0.0502797
25	0.0563891	15	0.0236717	25	0.0489033

В таблице 1.13 представлены результаты приближенного вычисления электрической центральности вершин графа с помощью метода последовательных приближений рекуррентной формулой (1.10) при $k = 2$ и при $k = 100$. Для сравнения приведены результаты первых 20 вершин.

Из таблицы 1.13 видно, что метод итераций при $k = 2$ позволил выявить 15 вершин из 20 имеющих наибольшее значение электрической центральности. Таким образом, метод последовательных приближений дает хорошую оценку рангов вершин графа уже после двух итераций.

При $k = 100$ метод итераций выявил все 20 первых вершин по значению электрической центральности, где первые 8 вершин получили верный ранг. Вершины 66 и 63, а также вершины 24 и 47 получили разные ранги в сравнении с истинными значениями электрической центральности, это можно объяснить близким значением центральности для данных вершин.

Таким образом, можно сделать вывод о том, что в некоторых прикладных задачах, где требуется выявить множество вершин имеющих наибольшее значение электрической центральности, применение стохастического метода Монте-Карло и метода последовательных приближений является оправданной и позволяет существенно уменьшить вычислительную сложность исходного алгоритма.

Глава 2

Теоретико-игровые модели центральности вершин в коммуникационных графах

2.1. Определение характеристической функции в кооперативной игре с помощью электрической центральности

Для определения центральности вершин в графе можно использовать теоретико-игровой подход. Предположим, что множество вершин графа является множеством игроков в некоторой кооперативной игре. Пусть $N = \{1, 2, \dots, n\}$ множество игроков. Через 2^N обозначим множество всех его подмножеств.

Кооперативной игрой n лиц будем называть пару $\langle N; w \rangle$, где $N = \{1, 2, \dots, n\}$ — множество игроков, а $w: 2^N \rightarrow R$ отображение, предписывающее каждой коалиции $S \in 2^N$ некоторое численное значение такое, что $w(\emptyset) = 0$. Функция w называется характеристической функцией кооперативной игры [60].

Неориентированный граф $g = (N, E)$ состоит из множества вершин N и множества рёбер E . Рёбра представляют неупорядоченные пары вершин, которые мы будем обозначать ij , и $ij \in E$ означает, что вершины $i \in N$ и $j \in N$ связаны в графе g .

Вершины графа идентифицируются с игроками, а ребро графа ij означает, что игроки i и j могут взаимодействовать напрямую, если и только если $ij \in g$.

Для графа g последовательность различных вершин $\{i_1, \dots, i_k\}$, $k \geq 2$ есть путь от i_1 до i_k , если для всех $h = 1, \dots, k-1$, $i_h i_{h+1} \in g$. Длина пути l — число рёбер в нём, $l = k-1$. Расстоянием между двумя вершинами является длина минимального пути между этими вершинами.

Граф g на множестве игроков N связный, если для любых двух вершин существует путь в g от одной вершины к другой.

Коалиция S связна, если любая пара игроков в S связана путём, состоящим

только из игроков данной коалиции.

Любая коалиция $S \subseteq N$ единственным образом разбивается графом g на максимальные связные подкоалиции, называемые компонентами связности. Множество всех связных компонент коалиции S обозначим через $S|g$.

$$S|g = \{\{i|i \text{ и } j \text{ связаны в } S \text{ через } g\}|j \in S\}.$$

$S|g$ можно интерпретировать как набор более мелких коалиций, на которые разбита коалиция S , если игроки могут взаимодействовать только в соответствии со связями в графе g .

$N|g$ означает множество всех связных компонент в g .

Граф, получаемый удалением из существующего графа g связи ij мы будем обозначать $g - ij$, а граф, получаемый добавлением в граф g связи ij обозначим $g + ij$.

Ребро ij является мостом в графе g , если $g - ij$ имеет больше компонент, чем g .

Пусть дана кооперативная игра w с множеством игроков N и граф g , вершинами которого являются игроки. Для каждого игрока i , данного графа g и характеристической функции w вектор Майерсона $Y(w, g) = (Y_1(w, g), \dots, Y_n(w, g))$ определяется следующими аксиомами:

A1 (Аксиома компонентной эффективности). Если S связная компонента, то сумма выигрышей игроков коалиции S равна ценности всей коалиции, т.е. $\forall S \in N|g$

$$\sum_{i \in S} Y_i(w, g) = w(S). \quad (2.1)$$

A2 (Аксиома справедливости). $\forall g, \forall ij \in g$ оба игрока одинаково получают выгоду или теряют от создания связи

$$Y_i(w, g) - Y_i(w, g - ij) = Y_j(w, g) - Y_j(w, g - ij). \quad (2.2)$$

Если для любой коалиции S определить характеристическую функцию как

$$w_g(S) = \sum_{K \in S|_g} w(K),$$

то вектор Майерсона может быть вычислен по формуле

$$Y_i(w, g) = \sum_{S \subset N \setminus \{i\}} (w_g(S \cup i) - w_g(S)) \frac{s!(n-s-1)!}{n!}, \quad (2.3)$$

где $s = |S|$, $n = |N|$.

Далее рассмотрим математическую модель кооперативной игры, где характеристическая функция определяется на основе модели электрической цепи.

Рассмотрим кооперативную игру, где характеристическая функция определяется суммой средних токов, протекающих через вершины соответствующей коалиции. При этом, немного модифицируем модель, исключив из нее входящие токи. Отметим, что слагаемое $\frac{1}{2}b_s(v)$ в формуле (1.11) дает одинаковый вклад для всех вершин коалиции, и не несет информации об относительном соотношении протекающих токов через вершины коалиции.

Таким образом, определим характеристическую функцию в кооперативной игре как

$$w(K) = \frac{1}{|K|} \sum_{s \in K} x_K^s(v), \quad K \subset V, \quad (2.4)$$

где

$$x_K^s(v) = \frac{1}{2} \sum_{e: v \in e} x_e^s.$$

В качестве меры центральности можно принять вектор Майерсона в кооперативной игре с характеристической функцией представленной (2.4). Назовем данную меру электрической центральностью по Майерсону.

Таким образом, при вычислении электрической центральности по Майерсону определяется как средний вклад вершины по всем возможным коалициям, где вклад вершины зависит от всех путей в графе, проходящих через нее.

Отметим, что кооперация игроков ограничена графом, т.е. игроки могут кооперироваться, только если они связаны в графе.

2.2. Использование вектора Майерсона для ранжирования вершин коммуникационного графа

В [61] предлагается использовать метод кооперативной теории игр для вычисления центральности вершин графа. В [61] была предложена процедура получения дележа, в которой используется производящая функция, а также доказано, что полученный в результате делёж совпадает с вектором Майерсона. В качестве характеристической функции использовалась специальная функция, которая зависит от числа геодезических путей в графе проходящих через вершину. Известно, что вычисление вектора Майерсона [62] требует больших вычислительных затрат, но вычислительная сложность алгоритма поиска вектора Майерсона в данной кооперативной игре равна $O(n^3)$. Центральность вершин графа определенную данным способом назовем центральностью на основе вектора Майерсона.

В качестве примера вычислим меру центральности на основе вектора Майерсона и электрическую центральность по Майерсону для взвешенного графа, представленного на рисунке 1.8. Напомним, что можно сделать интуитивный вывод о том, что вершина C должна иметь значение центральности большее, чем вершина B . Результаты вычислений меры центральности на основе вектора Майерсона для $r = 0.2$ и электрической центральности по Майерсону для $\delta = 0.5$ приведены в таблице 2.1. Результаты ранжирования подтверждают интуитивный вывод, сделанный ранее.

Далее рассмотрим пример графа в виде двух соединённых звёзд, представленного на рисунке 1.10. Результаты вычислений меры центральности на основе вектора Майерсона для $r = 0.2$ и электрической центральности по Майерсону для $\delta = 0.5$ без учёта весов рёбер приведены в таблице 2.2.

Таблица 2.1

Значения центральностей на основе вектора Майерсона для взвешенного графа, представленного на рис. 1.8.

Вершина	Электрическая центральность по Майерсону ($\delta = 0.5$)	Центральность по Майерсону ($r = 0.2$)
<i>A</i>	0.342324	0.5533
<i>B</i>	0.186156	0.3133
<i>C</i>	0.219573	0.54
<i>D</i>	0.120414	0.15339

Как и ожидалось в результате ранжирования обоими рассматриваемыми методами вершины 1 и 5 имеют максимальное значение центральности, а вершины 2, 3 и 4 получили значение центральности большее, чем вершины-листья: 6, 7, 8, 9, 10, 11, 12 и 13. Вершины 2 и 4 получили равное значение центральности. Отличие в результатах ранжирования заключается только в том, что для электрической центральности по Майерсону для $\delta = 0.5$ вершина 3 получила большую значимость, чем вершины 2 и 4. Более того, вершины 3, 2 и 4 получили равное значение центральности на основе вектора Майерсона для $r = 0.2$. Это объясняется тем, что данная мера центральности учитывает только геодезические пути. Примечательно, что вершина 3 получала меньшее значение центральности, чем вершины 2 и 4 как для электрической центральности по Майерсону при $\delta = 1$, так и для значения PageRank с $\alpha = 0.85$ (см. таблицу 1.7).

В таблице 2.3 приведены результаты вычислений мер центральности для двух соединённых звёзд представленной на рисунке 1.10 с учётом весов рёбер.

Результаты ранжирования полностью соответствуют интуитивному выводу о центральности вершин графа с учетом весов ребер. Вершины 1 и 5 получили высокое значение центральности относительно остальных вершин графа,

также вершина 5 получила большее значение центральности, чем вершина 1, т.к. у нее имеется более весомое ребро с вершиной 10. Вершина 4 стоит выше, чем вершина 2 за счет ребра с вершиной 3 с весом 4. Из вершин-листьев графа наибольшую центральность должна получить вершина 10, далее идут вершины 12, 7 и 9. При этом центральность вершины 12 выше центральности вершин 7 и 9 за счет близости к ребру с весом 3.

Таблица 2.2

Результаты расчетов центральности на основе вектора Майерсона для $r = 0.2$ и электрической центральности по Майерсону для $\delta = 0.5$ без учета весов ребер.

Вершина	Электрическая центральность по Майерсону ($\delta = 0.5$)	Вершина	Центральность по Майерсону ($r = 0.2$)
1	0.223643	1	0.65884
5	0.223643	5	0.65884
3	0.140689	2	0.29545
2	0.135323	4	0.29545
4	0.135323	3	0.26345
6	0.038015	6	0.15574
7	0.038015	7	0.15574
8	0.038015	8	0.15574
9	0.038015	9	0.15574
10	0.038015	10	0.15574
11	0.038015	11	0.15574
12	0.038015	12	0.15574
13	0.038015	13	0.15574

Однако, вершина 3 для электрической центральности по Майерсону при $\delta = 0.5$ получает большее значение значимости, чем вершина 4, в отличие от

результатов ранжирования центральностью на основе вектора Майерсона для $r = 0.2$, методом PageRank с $\alpha = 0.85$ и электрической центральности при $\delta = 1$ (см. таблицу 1.8).

Вершина 2 получила более высокий ранг, чем вершина 10 для электрической центральности по Майерсону при $\delta = 0.5$, в отличие от результатов ранжирования центральностью на основе вектора Майерсона для $r = 0.2$.

Таблица 2.3

Результаты расчетов центральности на основе вектора Майерсона для $r = 0.2$ и электрической центральности по Майерсону для $\delta = 0.5$ с учетом весов ребер.

Вершина	Электрическая центральность по Майерсону ($\delta = 0.5$)	Вершина	Центральность по Майерсону
5	0.253644	5	1.25188
1	0.232985	1	0.999936
3	0.170126	4	0.794229
4	0.1663	3	0.726229
2	0.144339	10	0.529138
10	0.0547762	2	0.396896
12	0.0492106	12	0.379426
7	0.0484886	7	0.34083
9	0.0484886	9	0.340832
11	0.0349946	11	0.203046
13	0.0349946	13	0.203046
6	0.0345538	6	0.183749
8	0.0345538	8	0.183749

2.3. Сравнение с другими моделями ранжирования

Проведем сравнительный анализ результатов ранжирования методами, рассматриваемыми в данной главе. В качестве примеров возьмем графы, представленные в [65]. Данные графы состоят из двух групп - полные графы, состоящие из 5 вершин и соединенных между собой посредством небольшого числа вершин.

На рисунке 2.1 представлен граф, демонстрирующий слабые места в работе методов ранжирования основанных на геодезических путях в графе. В данном графе две группы из 5 вершин соединены тремя вершинами 1, 2 и 3. Вершины 2 и 11 естественно получают более высокие значения центральности, т.к. все геодезические пути между двумя группами проходят через эти вершины. С другой стороны вершина 1 получит малое значение центральности методами, основанными на геодезических путях, проходящих через вершину 1. Однако в большинстве случаев в реальной жизни вершина 1 будет играть существенную роль в распространении информации. Естественно, что информация может передаться от одной вершины к другой через третьего общего знакомого, а не только напрямую непосредственно.

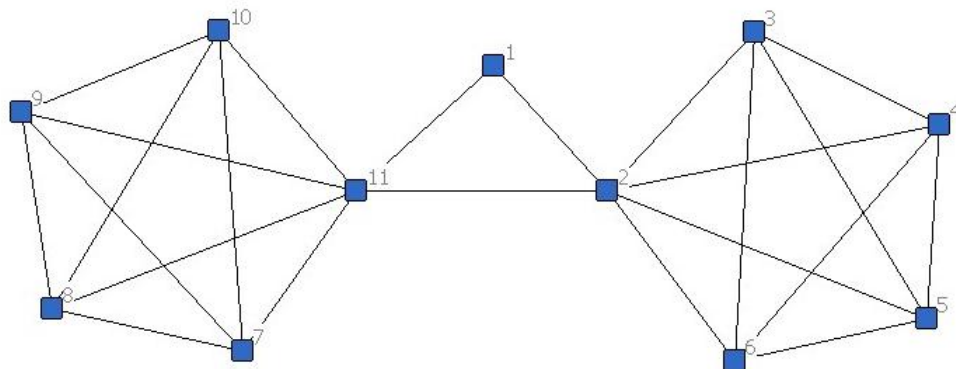


Рисунок 2.1. Пример невзвешенного графа из 11 вершин.

Результаты вычислений мер центральностей для графа, представленного на рисунке 2.1, приведены в таблице 2.4. Здесь мы видим, что вершина 1 получила наименьший ранг по центральности по Майерсону при $r = 0.2$, т.к. данный метод основан на геодезических путях. Остальные рассматриваемые методы дают вершине 1 адекватное третье место.

Результаты расчетов центральностей для графа, представленного на рисунке 2.1.

Вер- шина	Электр. центр-сть по Майерсону ($\delta = 0.5$)	Вер- шина	Электр. центр-сть ($\delta = 0.5$)	Вер- шина	Электр. центр-сть, описанная в [29]	Вер- шина	Центр-сть по Майерсону ($r = 0.2$)
2	0.291652	2	0.294799	2	0.819259	2	0.792
11	0.291652	11	0.294799	11	0.819259	11	0.792
1	0.0996015	1	0.147207	1	0.407407	3	0.43467
3	0.0716545	3	0.127417	3	0.328889	4	0.43467
4	0.0716545	4	0.127417	4	0.328889	5	0.43467
5	0.0716545	5	0.127417	5	0.328889	6	0.43467
6	0.0716545	6	0.127417	6	0.328889	7	0.43467
7	0.0716545	7	0.127417	7	0.328889	8	0.43467
8	0.0716545	8	0.127417	8	0.328889	9	0.43467
9	0.0716545	9	0.127417	9	0.328889	10	0.43467
10	0.0716545	10	0.127417	10	0.328889	1	0.30667

На рисунке 2.2 представлен граф, который приводился в [65] в качестве примера для демонстрации некорректной работы меры центральности, основанной на потоках [44].

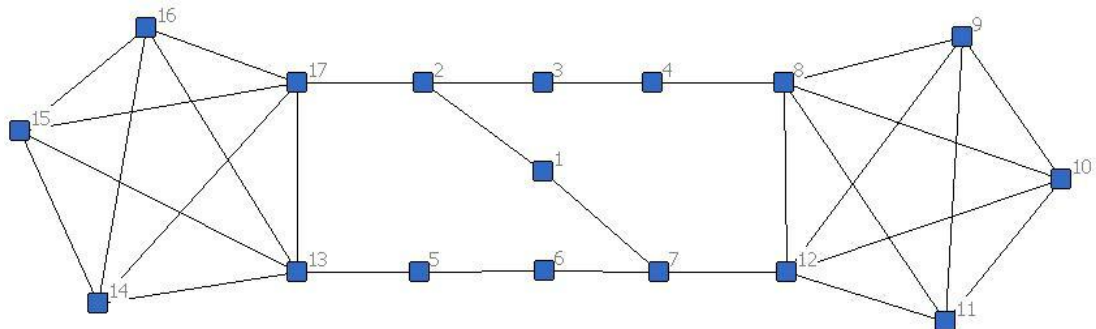


Рисунок 2.2. Пример невзвешенного графа из 17 вершин.

Для дальнейшего анализа центральности вершин графа, представленного на рисунке 2.2, отметим, что в [65] указано, что ранг вершины 1 должно быть больше, чем ранги вершин 9-11 и 14-16, а ранг вершин 2 и 7 должен быть выше ранга вершины 1. Также отметим, что при расчете меры центральности предложенной в [44] учитываются не только геодезические пути.

Вычислим электрическую центральность по Майерсону для графа, представленного на рисунке 2.2. Для сравнения приведём результаты расчётов метода PageRank, электрической центральности по методу, описанному в [29] и центральности по Майерсону с параметром $r = 0.5$.

Результаты вычислений мер центральностей для графа, представленного на рисунке 2.2, приведены в таблице 2.5.

Таблица 2.5

Результаты расчетов мер центральностей для графа, представленного на рисунке 2.2.

Вер- шина	Электр. центр-сть по Майерсону ($\delta = 0.5$)	Вер- шина	Электр. центр-сть ($\delta = 0.5$)	Вер- шина	Электр. центр-сть, описанная в [29]	Вер- шина	Центр-ть по Майер- сону $r = 0.2$
2	0.144037	2	0.156429	2	0.534253	12	2.44643
7	0.144037	7	0.156429	7	0.534253	17	2.44643
12	0.13073	12	0.148002	12	0.472929	2	2.3256
17	0.13073	17	0.148002	17	0.472929	7	2.3256
8	0.112156	8	0.137507	8	0.409149	8	2.14196
13	0.112156	13	0.137507	13	0.409149	13	2.14196
4	0.082407	4	0.108608	3	0.364358	1	1.50551
5	0.082407	5	0.108608	6	0.364358	4	1.42946
1	0.0727389	3	0.107291	4	0.358045	5	1.42946
6	0.0701525	6	0.107291	5	0.358045	3	1.39509
3	0.0701525	1	0.103644	1	0.302778	6	1.39509

9	0.0455176	9	0.0842484	9	0.220014	9	1.33363
10	0.0455176	10	0.0842484	10	0.220014	10	1.33363
11	0.0455176	11	0.0842484	11	0.220014	11	1.33363
14	0.0455176	14	0.0842484	14	0.220014	14	1.33363
15	0.0455176	15	0.0842484	15	0.220014	15	1.33363
16	0.0455176	16	0.0842484	16	0.220014	16	1.33363

Как и ожидалось вершины *2*, *7*, *12*, *17*, *8* и *13* получили максимальные ранги. Примечательно, что центральность по Майерсону при $r = 2$, основанная на геодезических путях, дает самое высокое 7-е место вершине *1* по сравнению с остальными методами. Электрическая центральность оп Майерсону при $\delta = 0.5$ дает вершине *1* 9-е место. Электрические центральности дают вершине *1* ранг выше, чем ранги вершин *9-11* и *14-16*. Таким образом, на данном примере все рассматриваемые методы ранжирования имеют адекватные результаты.

Методы электрической центральности при $\delta = 0.5$ и электрической центральности, описанной в [29] дают близкие результаты ранжирования вершин. Это объясняется тем, что оба эти метода основаны на модели электрических цепей и в расчете центральности используются правила Кирхгофа.

Разница в рангах вершин *3*, *6*, *4* и *5* является более чувствительной, и для всех рассмотренных методов данные вершины получили разные результаты.

Глава 3

Компьютерное моделирование в задачах анализа коммуникационных сетей

Анализ социальных сетей (social network analysis) исследует структуру взаимодействия между социальными сущностями. Этими социальными сущностями могут выступать люди, организации, города, страны, веб-страницы или научные публикации [76].

В данной главе рассматриваются примеры применения предложенного метода ранжирования вершин графов коммуникационных сетей из реальной жизни. Представим результаты ранжирования электрической центральностью на графах транспортных сетей Китая, Финляндии и Транссибирской магистрали, а также на графе публикаций Math-Net.ru написанных в соавторстве. Также представим пример ранжирования сайтов научных учреждений Российской академии наук и пример применения электрической меры близости в онтологической модели компетенций. Для вычисления электрической центральности вершин разработан программный комплекс в пакете Mathematica.

3.1. Приложения для транспортных сетей

Задача поиска значимости вершин в графе может возникнуть для транспортных сетей. В [7] в качестве примеров для поиска значимости вершин в графе вектором Майерсона рассматривались графы Транссибирской железнодорожной магистрали, а также транспортных сетей Китая и Финляндии. Используем данные примеры для анализа работы электрической центральности.

3.1.1. Транссибирская железнодорожная магистраль

Как мы определили в примере 1.4, если рассматривать евразийский континент, то географическое положение России предоставляет ей преимущество, как пе-

ред европейскими, так и перед азиатскими странами. Ключевую роль в реализации потенциальных преимуществ географического положения России играет Транссибирская магистраль.

„Транссиб соединяет Европейскую часть, Урал, Сибирь и Дальний Восток России, а говоря уже — российские западные, северные и южные порты, а также железнодорожные выходы в Европу (Санкт-Петербург, Мурманск, Новороссийск), с одной стороны, с тихоокеанскими портами и железнодорожными выходами в Азию (Владивосток, Находка, Забайкальск). Пропускная способность: 100 млн. тонн грузов в год.“¹

На рисунке 3.1 представлена схема перевозок грузов с участием Транссибирской магистрали.

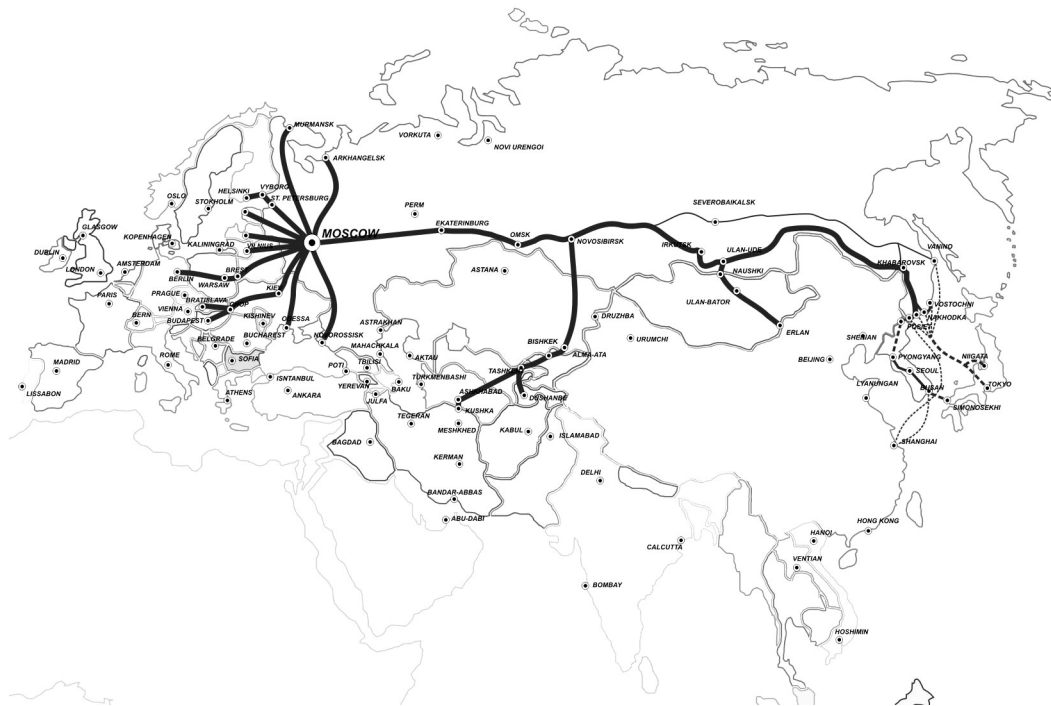


Рисунок 3.1. Схема перевозки грузов с участием Транссибирской магистрали.

На рисунке 3.2 представлен граф, построенный на схеме перевозок грузов с участием Транссибирской магистрали. Здесь мы видим, что граф имеет древовидную структуру и соответственно не имеет циклов.

¹<https://ru.wikipedia.org>

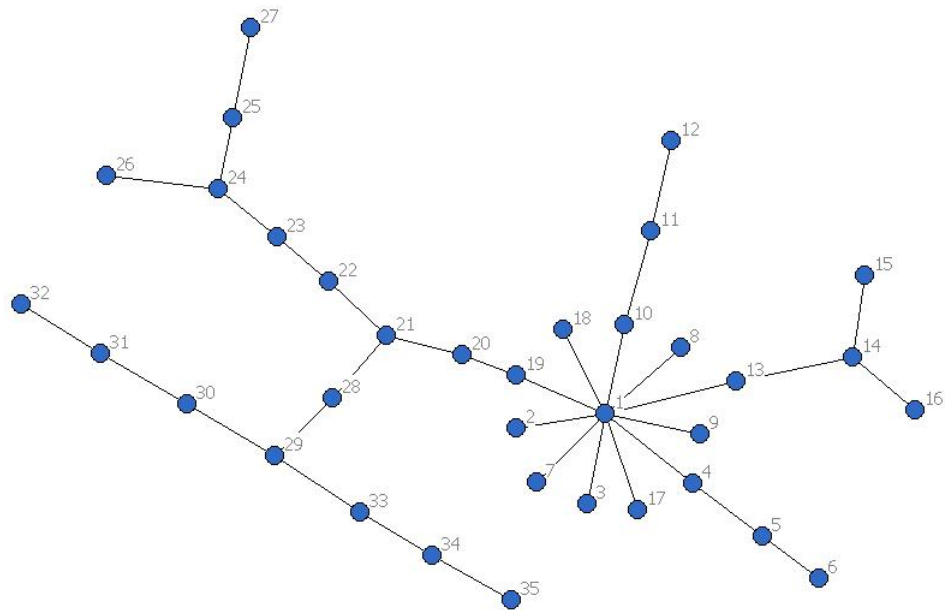


Рисунок 3.2. Граф, соответствующий схеме перевозки грузов с участием Транссибирской магистрали.

Из рисунков 3.1 и 3.2 можно выделить интуитивно следующие вершины: Москва, Новосибирск и Улан-Удэ. Также относительно высокие ранги должны получить Екатеринбург, Омск и Иркутск, так как они соединяют центры звезд.

В таблице 3.1 представлены результаты расчетов электрической центральности с параметрами $\delta = 1$, $\delta = 0.5$, $\delta = 0.3$ и $\delta = 0.1$.

Таблица 3.1

Результаты расчетов электрической центральности графа Транссибирской магистрали.

$\delta = 1$			$\delta = 0.5$			$\delta = 0.1$		
1	Москва	0.1543	1	Москва	0.21251	1	Москва	0.34192
21	Новосибирск	0.0562	21	Новосибирск	0.08055	21	Новосибирск	0.20317
29	Улан-Удэ	0.0559	29	Улан-Удэ	0.07861	19	Екатеринбург	0.17403
24	Ташкент	0.0530	24	Ташкент	0.06996	29	Улан-Удэ	0.16297
14	Чор	0.0515	14	Чор	0.06676	20	Омск	0.15738
19	Екатеринбург	0.0457	19	Екатеринбург	0.06542	28	Иркутск	0.13533
13	Киев	0.0444	13	Киев	0.06168	22	Алма-Ата	0.12211
4	С. Петербург	0.0431	20	Омск	0.05879	24	Ташкент	0.11852
10	Брест	0.0431	28	Иркутск	0.05867	13	Киев	0.11813
28	Иркутск	0.0423	4	С. Петербург	0.05826	23	Бишкек	0.11218
20	Омск	0.0419	10	Брест	0.05826	14	Чор	0.10584
22	Алма-Ата	0.0412	22	Алма-Ата	0.05609	4	С. Петербург	0.10120

23	Бишкек	0.0410	23	Бишкек	0.05515	10	Брест	0.10120
30	Наушки	0.0407	30	Наушки	0.05377	30	Наушки	0.09370
33	Чита	0.0407	33	Чита	0.05377	33	Чита	0.09370
25	Ашхабад	0.0387	25	Ашхабад	0.04821	5	Выборг	0.07286
5	Выборг	0.0382	5	Выборг	0.04801	11	Варшава	0.07286
11	Варшава	0.0382	11	Варшава	0.04801	25	Ашхабад	0.07047
31	Улан-Батор	0.0379	31	Улан-Батор	0.04705	31	Улан-Батор	0.07033
34	Хабаровск	0.0379	34	Хабаровск	0.04705	34	Хабаровск	0.07033
2	Архангельск	0.0274	2	Архангельск	0.03186	2	Архангельск	0.03856
3	Мурманск	0.0274	3	Мурманск	0.03186	3	Мурманск	0.03856
7	Таллин	0.0274	7	Таллин	0.03186	7	Таллин	0.03856
8	Рига	0.0274	8	Рига	0.03186	8	Рига	0.03856
9	Вильнюс	0.0274	9	Вильнюс	0.03186	9	Вильнюс	0.03856
17	Одесса	0.0274	17	Одесса	0.03186	17	Одесса	0.03856
18	Новороссийск	0.0274	18	Новороссийск	0.03186	18	Новороссийск	0.03856
26	Душанбе	0.0259	26	Душанбе	0.02979	15	Братислава	0.03674
15	Братислава	0.0259	15	Братислава	0.02966	16	Будапешт	0.03674
16	Будапешт	0.0259	16	Будапешт	0.02966	26	Душанбе	0.03669
27	Кушка	0.0253	27	Кушка	0.02882	6	Хельсинки	0.03620
6	Хельсинки	0.0252	6	Хельсинки	0.02879	12	Берлин	0.03620
12	Берлин	0.0252	12	Берлин	0.02879	27	Кушка	0.03576
32	Эрлан	0.0252	32	Эрлан	0.02866	32	Эрлан	0.03574
35	Находка	0.0252	35	Находка	0.02866	35	Находка	0.03574

Как и предполагалось ранее Москва, Новосибирск и Улан-Удэ получили наибольшие значения центральности. Интересно отметить, что при уменьшении δ все большее значение электрической центральности получают такие вершины как Екатеринбург, Омск и Иркутск, другими словами вершины соединяющие центры локальных звезд. Это объясняется тем, что при уменьшении δ ток все больше „растекается“ по сети, давая преимущество вершинам подобным Екатеринбургу, Омску и Иркутску. Соответственно с другой стороны при увеличении δ преимущество центров локальных звезд становится все очевиднее.

3.1.2. Железные дороги Финляндии

Далее рассмотрим граф железных дорог Финляндии. Данный граф уже рассматривался для демонстрации работы классической центральности в примере 1.2. Граф является неориентированным, невзвешенным и имеет циклы.

В таблице 3.2. представлены результаты вычисления электрической центральности при $\delta = 1$, $\delta = 0.5$, $\delta = 0.3$ и $\delta = 0.1$. Представлены первые 20 значений центральностей вершин.

Таблица 3.2

Результаты расчетов электрической центральности вершин графа железных дорог Финляндии.

$\delta = 1$		$\delta = 0.5$		$\delta = 0.3$		$\delta = 0.1$	
Pieksämäki	0.0553879	Pieksämäki	0.0806424	Tampere	0.110429	Tampere	0.226
Kouvola	0.0537417	Tampere	0.0793138	Pieksämäki	0.104781	Pieksämäki	0.173729
Tampere	0.0511364	Kouvola	0.0755048	Kouvola	0.0948421	Kouvola	0.146008
Seinäjoki	0.0444503	Seinäjoki	0.0625113	Seinäjoki	0.0793459	Toijala	0.13026
Oulu	0.0440442	Oulu	0.0621804	Oulu	0.07846	Seinäjoki	0.127561
Toijala	0.0372564	Toijala	0.054374	Toijala	0.0721898	Oulu	0.117393
Ylivieska	0.0360977	Ylivieska	0.0506325	Jyväskylä	0.0653608	Jyväskylä	0.113961
Isalmi	0.0359229	Jyväskylä	0.0503658	Ylivieska	0.0643933	Lahti	0.106321
Haapa-mäki	0.0359156	Isalmi	0.0500001	Lahti	0.0641123	Ylivieska	0.102321
Jyväskylä	0.0355945	Lahti	0.0499328	Isalmi	0.0630898	Haapa-mäki	0.10217
Lahti	0.0355635	Haapa-mäki	0.0496914	Haapa-mäki	0.0629075	Isalmi	0.0980761
Pännäinen	0.0353323	Pännäinen	0.0487965	Pännäinen	0.061463	Pännäinen	0.0979615
Karjaa	0.0352976	Karjaa	0.0486962	Karjaa	0.0607612	Karjaa	0.0901857
Turku	0.0339459	Joensuu	0.0454316	Joensuu	0.05573	Riihimäki	0.083743
Joensuu	0.0339186	Turku	0.0454087	Turku	0.0554946	Tikkurila	0.0835334
Parikkala	0.0335795	Parikkala	0.0445712	Tikkurila	0.0550157	Joensuu	0.082955
Tikkurila	0.0331973	Tikkurila	0.0445199	Riihimäki	0.0546213	Turku	0.0813988
Riihimäki	0.0331601	Riihimäki	0.0443363	Parikkala	0.0542248	Parkano	0.0796735
Vihanti	0.0321743	Vihanti	0.0422574	Kontiomäki	0.0518454	Kuopio	0.0795255
Kontiomäki	0.0313966	Kontiomäki	0.0420665	Vihanti	0.0508724	Parikkala	0.0790207

При относительно больших значениях, т.е. при $\delta = 1$ первые три места получили те же города что и при использовании классической центральности c_B : Пиексямяки, Коуволла и Тампере. При уменьшении параметра δ , Тампере

получает все больший ранг и при $\delta = 0.3$ и $\delta = 0.1$ уже получает наибольшее значение центральности. Подобное устойчивое увеличение ранга наблюдается и у городов Тоижала, Ювяскила, Лахти и Риихимяки.

3.1.3. Железные дороги Китая

Известно, что Китай обладает крупнейшей в мире сетью скоростных железных дорог. На рисунке 3.3 представлена схема высокоскоростных железных дорог Китая. Скорость движения поездов по данной схеме составляет до 350 км/ч. На некоторых участках данной схемы работает высокоскоростной поезд на магнитной левитации, скорость движения которого достигает 431 км/ч. Граф составленный на основе схемы высокоскоростных железных дорог Китая состоит из 49 вершин и 77 ребер, диаметр графа равен 10, средняя длина пути — 4.44, число геодезических путей — 2352.



Рисунок 3.3. Схема высокоскоростных железных дорог Китая.

В таблице 3.3. представлены результаты вычисления электрической центральности при $\delta = 1$, $\delta = 0.5$, $\delta = 0.3$ и $\delta = 0.1$. Представлены первые 20

значений центральностей вершин.

Таблица 3.3

Результаты расчетов электрической центральности вершин графа
высокоскоростных железных дорог Китая.

$\delta = 1$		$\delta = 0.5$		$\delta = 0.3$		$\delta = 0.1$	
Ухань	0.06904	Ухань	0.10690	Ухань	0.14688	Ухань	0.28357
Гуанчжоу	0.05650	Чэнду	0.07834	Наньчан	0.10579	Наньчан	0.21902
Чэнду	0.05544	Наньчан	0.07705	Сиань	0.10117	Сиань	0.17551
Чанша	0.05442	Сиань	0.07663	Чэнду	0.09963	Чжэнчжоу	0.16464
Сиань	0.05230	Гуанчжоу	0.07602	Чанша	0.09638	Чанша	0.15927
Наньчан	0.05141	Чанша	0.07589	Чжэнчжоу	0.09457	Чэнду	0.15923
Тайюань	0.05039	Чжэнчжоу	0.07237	Гуанчжоу	0.09120	Тайюань	0.13446
Чжэнчжоу	0.05024	Тайюань	0.07030	Тайюань	0.08837	Шицзячжуан	0.13258
Гуйян	0.05002	Гуйян	0.06795	Гуйян	0.08397	Гуйян	0.12678
Шэньян	0.04865	Шэньян	0.06690	Шицзячжуан	0.08344	Шэньян	0.12191
Шицзячжуан	0.04753	Шицзячжуан	0.06602	Шэньян	0.08295	Гуанчжоу	0.12159
Цзинань	0.04717	Цзинань	0.06395	Пекин	0.07926	Пекин	0.11973
Пекин	0.04670	Пекин	0.06377	Цзинань	0.07869	Цзинань	0.11732
Ханчжоу	0.04564	Ханчжоу	0.06043	Фучжоу	0.07337	Хуаншань	0.11335
Фучжоу	0.04462	Фучжоу	0.05983	Ханчжоу	0.07254	Сюйчжоу	0.11215
Хэфэй	0.04385	Хэфэй	0.05873	Хэфэй	0.07201	Фучжоу	0.11200
Хуаншань	0.04351	Хуаншань	0.05810	Хуаншань	0.07152	Хэфэй	0.11013
Гуйлинь	0.04260	Гуйлинь	0.05595	Сюйчжоу	0.06840	Ханчжоу	0.10156
Чунцин	0.04198	Чунцин	0.05528	Гуйлинь	0.06695	Бэнбу	0.09875
Нанкин	0.03905	Сюйчжоу	0.05386	Чунцин	0.06612	Гуйлинь	0.09265

Наиболее крупные города, такие как Ухань, Сиань, Чэнду, Гуанчжоу и Пекин получили высокие ранги.

Примечательно, что при уменьшении параметра δ , такие города как Сиань, Наньчан, Чжэнчжоу, Шицзячжуан, Пекин и Хуншань получают все больший ранг и при этом остальные города сохраняют свои ранги относительно друг друга.

3.1.4. Метро города Москвы

В качестве примера рассмотрим самую знаменитую и узнаваемую транспортную сеть в России: схему московского метрополитена. По интенсивности использования метро Москвы занимает 4-е место после метрополитенов Пекина, Токио и Сеула. Московское метро также известно богатой отделкой ряда станций образцами искусства эпохи социалистического реализма². Московский метрополитен состоит из 12 линий общей протяженностью 327.5 км в двухпутном исчислении³. В московском метро 196 станций, 44 из которых признаны объектами культурного наследия. Дневной пассажиропоток по данным 2014 года составляет 6,72 млн. человек⁴.

На рисунке 3.4 представлена схема московского метрополитена.

²"Московский метрополитен."Википедия, свободная энциклопедия. 27 авг 2015, 10:09 UTC. 2 сен 2015, 07:18 <<http://ru.wikipedia.org/?oldid=72981902>>.

³Метрополитен в цифрах. Официальный сайт Московского метрополитена. Проверено 29 декабря 2011. Архивировано из первоисточника 5 января 2013

⁴Основные технико-эксплуатационные характеристики метрополитенов за 2014 год. Международная ассоциация «Метро».



Рисунок 3.4. Схема линий московского метрополитена.

При построении графа на основе схемы московского метрополитена были сделаны следующие допущения:

- станции метро с пересадками были объединены между собой в одну вершину (см. Таб. 3.4);
- ребро инцидентное вершинам соответствующим станциям Охотный ряд, Театральная и Площадь революции с одной стороны, и станциями Арбатская, Боровицкая, Александровский сад и Библиотека им. Ленина с другой стороны, имеет вес равный 2, т.к. они связаны двумя линиями Сокольнической и Арбатско-Покровской.

Итоговый граф, построенный на основе схемы московского метрополитена, имеет следующие характеристики: состоит из 134 вершин и 157 ребер, диаметр графа равен 24, средняя длина пути — 9.37 и содержит 17822 геодезических путей.

Таблица 3.4

Соответствие станций метро с пересадками с вершинами графа метро Москвы.

Станции метро с пересадками (линия метро)	Название вершины графа
Арбатская, Библиотека им. Ленина, Боровицкая и Александровский сад	Арбатская
Киевская (Филевская, Арбатско-Покровская и Кольцевая)	Киевская
Чеховская, Пушкинская, Тверская	Тверская
Таганская (Кольцевая, Таганско-Краснопресненская), Марксистская	Таганская
Курская (Кольцевая, Арбатско-Покровская), Чкаловская	Курская
Охотный ряд, Театральная, Площадь революции	Охотный ряд
Третьяковская (Калининская, Калужско-Рижская), Новокузнецкая	Третьяковская
Краснопресненская, Баррикадная	Краснопресненская
Новослободская, Менделеевская	Новослободская
Проспект мира (Кольцевая, Калужско-Рижская)	Проспект мира
Комсомольская (Кольцевая, Сокольническая)	Комсомольская
Добрынинская, Серпуховская	Добрынинская
Белорусская (Кольцевая, Замоскворецкая)	Белорусская
Крестьянская застава, Пролетарская	Пролетарская
Павелецкая (Кольцевая, Замоскворецкая)	Павелецкая
Тургеневская, Чистые пруды	Чистые пруды
Октябрьская (Кольцевая, Калужско-Рижская)	Октябрьская
Парк культуры (Кольцевая, Сокольническая)	Парк культуры
Китай-город (Таганско-Краснопресненская, Калужско-Рижская)	Китай-город
Кузнецкий мост, Лубянка	Кузнецкий мост
Каширская (Каховская, Замоскворецкая)	Каширская
Севастопольская, Каховская	Севастопольская
Римская, Площадь Ильича	Римская

В таблице 3.5. представлены результаты вычисления электрической центральности вершин графа московского метрополитена при $\delta = 1$, $\delta = 0.5$ и $\delta = 0.1$. Представлены первые 20 значений центральностей вершин.

Таблица 3.5

Результаты расчетов электрической центральности вершин графа московского метрополитена.

$\delta = 1$		$\delta = 0.5$		$\delta = 0.1$	
Арбатская	0.0263452	Киевская	0.0371577	Таганская	0.0819931
Киевская	0.0260231	Арбатская	0.0369503	Курская	0.0791802
Тверская	0.0234674	Таганская	0.0337777	Киевская	0.0772832

Таганская	0.0230184	Тверская	0.0332158	Арбатская	0.074059
Курская	0.0223586	Курская	0.033053	Тверская	0.0701974
Охотный ряд	0.0217537	Охотный ряд	0.0308565	Охотный ряд	0.0667756
Третьяковская	0.0201127	Третьяковская	0.0288714	Краснопресненская	0.0661246
Краснопресненская	0.0191524	Краснопресненская	0.0281158	Третьяковская	0.0659292
Новослободская	0.0188647	Новослободская	0.0274504	Пролетарская	0.0641977
Проспект мира	0.0187075	Проспект мира	0.026965	Новослободская	0.0640436
Комсомольская	0.0186812	Добрынинская	0.0269042	Проспект мира	0.0613292
Добрынинская	0.018679	Комсомольская	0.0268122	Павелецкая	0.0613226
Белорусская	0.0181898	Пролетарская	0.0265051	Добрынинская	0.0604541
Пролетарская	0.0181622	Белорусская	0.0260404	Комсомольская	0.0598493
Павелецкая	0.0178707	Павелецкая	0.0259455	Белорусская	0.0588882
Чистые пруды	0.016937	Октябрьская	0.0236397	Октябрьская	0.0520483
Октябрьская	0.016803	Парк культуры	0.0232587	Парк культуры	0.0489979
Парк культуры	0.0167334	Чистые пруды	0.0232411	Чистые пруды	0.0470992
Китай-город	0.0153205	Китай-город	0.0208967	Каширская	0.0469468
Кузнецкий мост	0.015192	Каширская	0.0206601	Севастопольская	0.0441881

3.1.5. Граф стран Евразии

Представим результаты ранжирования вершин с помощью электрической центральности для графа стран Евразии. В таблице 3.6. представлены результаты вычисления электрической центральности при $\delta = 1$, $\delta = 0.5$, $\delta = 0.3$ и $\delta = 0.1$. Представлены первые 20 значений центральностей вершин.

Таблица 3.6

Результаты расчетов электрической центральности вершин графа стран Евразии.

$\delta = 1$		$\delta = 0.5$		$\delta = 0.3$		$\delta = 0.1$	
Россия	0.08932	Россия	0.13133	Россия	0.16832	Россия	0.25420
Китай	0.07648	Китай	0.10841	Китай	0.13557	Китай	0.19705
ФРГ	0.05049	Турция	0.07309	Турция	0.09568	Турция	0.15273
Турция	0.04933	ФРГ	0.06887	ФРГ	0.08408	Иран	0.12070
Франция	0.04757	Франция	0.06480	Франция	0.07825	ФРГ	0.11909
Австрия	0.04261	Иран	0.06081	Иран	0.07799	Польша	0.11893
Иран	0.04237	Польша	0.05797	Польша	0.07481	Ирак	0.10515
Сауд. Аравия	0.04157	Австрия	0.05784	Австрия	0.07080	Франция	0.10466
Польша	0.04030	Украина	0.05608	Украина	0.07014	Украина	0.10314

Украина	0.04022	Сауд. Аравия	0.05550	Ирак	0.06820	Австрия	0.10097
Сербия	0.03940	Венгрия	0.05385	Венгрия	0.06678	Венгрия	0.09680
Венгрия	0.03905	Ирак	0.05332	Сауд. Аравия	0.06674	Сауд.Аравия	0.09053
Индия	0.03828	Сербия	0.05164	Сербия	0.06162	Сербия	0.08435
Италия	0.03799	Италия	0.05121	Италия	0.06154	Италия	0.08224
Ирак	0.03731	Индия	0.04972	Индия	0.05875	Индия	0.07808
Афганистан	0.03530	Афганистан	0.04570	Афганистан	0.05411	Болгария	0.07313
Мьянма	0.03282	Мьянма	0.04305	Мьянма	0.05147	Афганистан	0.07282
Румыния	0.02959	Болгария	0.03999	Болгария	0.04941	Мьянма	0.07036
Болгария	0.02956	Румыния	0.03912	Румыния	0.04695	Таиланд	0.06609
Сирия	0.02890	Сирия	0.03837	Сирия	0.04655	Сирия	0.06540

По всем рассматриваемым параметрам δ Россия и Китай получают наибольшую центральность. В сравнении с результатами ранжирования методом PageRank Турция получает высокий ранг, к тому же ранг Турции растет при уменьшении значения параметра δ . Высокий ранг сохраняет ФРГ. Такие страны как Иран, Польша и Ирак стремительно улучшают свой ранг с уменьшением параметра δ .

3.2. Анализ портала математических публикаций Math-Net.ru

Проиллюстрируем работу метода электрической центральности для ранжирования вершин графа публикаций математического портала Math-Net.ru [3, 38].

На рис. 3.5 представлен фрагмент графа публикаций, составленного на основе данных математического портала Math-Net.ru. Общее число авторов математического портала Math-Net.ru на момент написания данной работы составляло 78839. Фрагмент графа публикаций Math-Net.ru, исследуемый в данной работе, содержит 7606 авторов и 10747 статей, написанных в соавторстве. Здесь вершины графа - это авторы статей, а вес ребра - это число совместных научных статей авторов. Отметим, что при построении данного графа не учитывались статьи, имеющие более 6 соавторов. Рис. 3.5 подготовлен в программном комплексе Gephi [20].

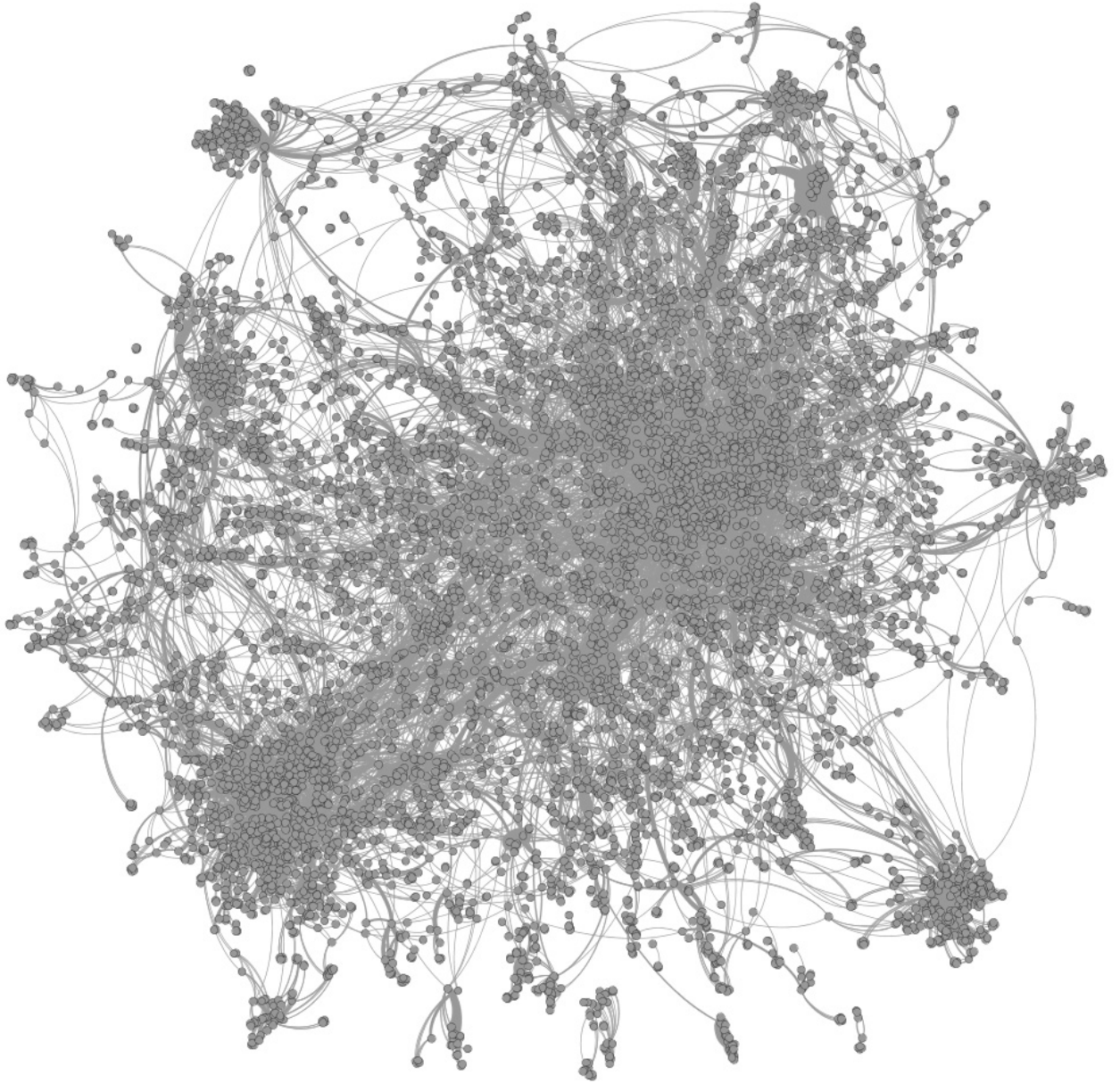


Рисунок 3.5. Фрагмент графа публикаций Math-Net.ru

Для большей наглядности результатов ранжирования, на рис. 3.6 представлена главная компонента графа, полученная путем удаления ребер с весом меньше 7 из графа, представленного на рис. 3.5.

На рис. 3.6 видно, что вершины 40, 34, 56 и 20 являются центрами „локальных“ звезд и, соответственно, должны иметь высокую центральность. Заметим, что вершина 32 также должна иметь высокую центральность, т.к. она соединяет в единый граф две отдельные компоненты.

В табл. 3.7 приведены результаты ранжирования для 35 первых вершин графа, представленного на рис. 3.6. Здесь в качестве методов ранжирования

используются электрическая центральность с параметром $\delta = 1$, PageRank с параметром $\alpha = 0.85$ и электрическая центральность (CF-betweenness), описанная в [29].

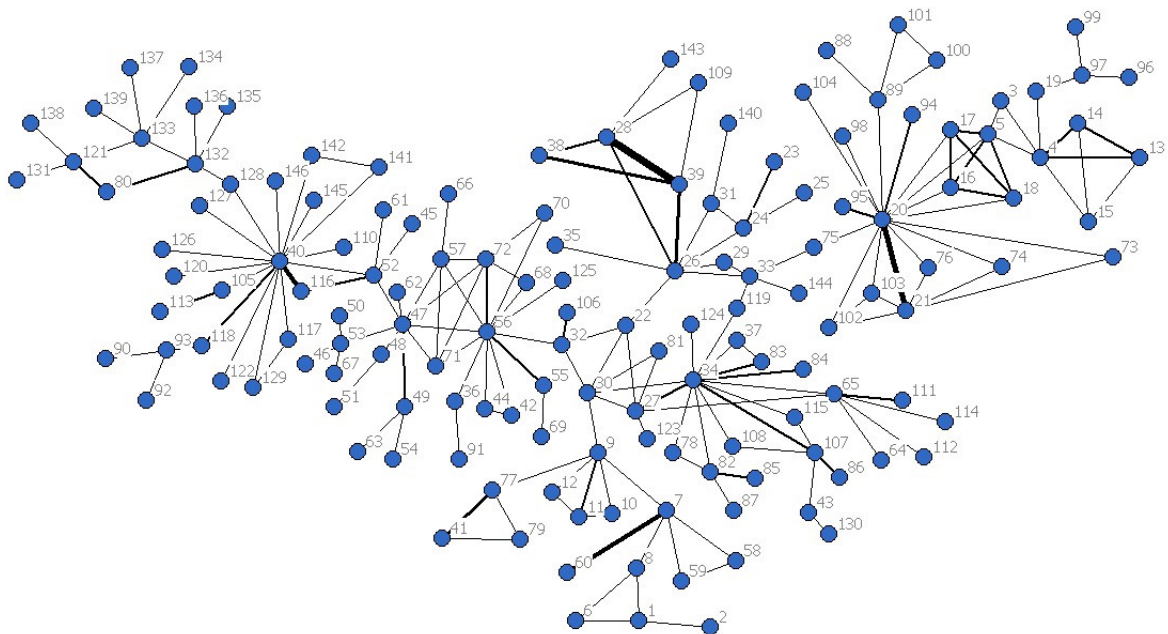


Рисунок 3.6. Главная компонента графа публикаций Math-Net.ru

Таблица 3.7

Результаты ранжирования вершин графа публикаций Math-Net.ru.

Вершина	Центральность (CF_B) $\delta = 1$	Вершина	PageRank	Вершина	CF-betweenness центральность
40	0.15740	40	0.04438	56	0.54237
34	0.14981	34	0.03285	32	0.53027
20	0.13690	20	0.03210	47	0.48222
47	0.12566	56	0.02774	22	0.41668
56	0.12518	47	0.02088	33	0.41361
26	0.10880	39	0.01874	34	0.39517
30	0.09098	28	0.01824	30	0.39426
9	0.08149	21	0.01695	52	0.37421
33	0.08024	65	0.01632	40	0.36946
32	0.07959	26	0.01552	26	0.35259
22	0.07903	107	0.01424	20	0.34413
52	0.07777	132	0.01399	75	0.32768
27	0.07510	116	0.01385	27	0.25769
5	0.06298	133	0.01365	116	0.24321
132	0.06001	72	0.01338	119	0.22100

4	0.05941	7	0.01313	72	0.20463
75	0.05711	27	0.01266	9	0.19517
116	0.05366	121	0.01249	128	0.15915
65	0.05296	53	0.01218	132	0.14554
7	0.05254	52	0.01173	5	0.14038
128	0.05104	9	0.01150	57	0.12240
72	0.04635	49	0.01115	4	0.11265
107	0.04411	4	0.01086	71	0.11243
119	0.04303	5	0.01071	7	0.10778
53	0.03878	97	0.01067	65	0.10217
133	0.03815	24	0.01046	29	0.07724
118	0.03615	82	0.01035	127	0.06892
24	0.03529	80	0.01012	133	0.06786
89	0.03499	30	0.01010	107	0.06504
57	0.03430	57	0.00993	121	0.06026
121	0.03304	118	0.00933	80	0.05988
28	0.03135	77	0.00929	53	0.05498
82	0.03126	89	0.00916	89	0.05492
21	0.03112	32	0.00912	8	0.05476
49	0.03102	17	0.00908	19	0.05469

Как и предполагалось ранее, вершины *40*, *34*, *56* и *20* получили высокие ранги по всем рассматриваемым методам ранжирования. Однако, для метода PageRank вершина *32* получила относительно низкий ранг (34 место).

В качестве примера, раскроем только некоторые данные по вершинам графа публикаций Math-Net.ru. Вершина *40* - Гельфанд И.М., *34* - Олейник О.А., вершина *56* - Кутателадзе С.С., а вершина *32* - Новиков С.П.

Таблица 3.8

Результаты применения метода итераций и метода Монте-Карло для ранжирования вершин графа публикаций Math-Net.ru.

Вершина	Центральность (CF_B) $\delta = 1$	Вершина	Метод итераций ($k = 2$)	Вершина	Метод Монте-Карло 10% вершин
40	0.15740	40	0.0762544	40	0.230855
34	0.14981	34	0.0541052	47	0.183219
20	0.13690	20	0.0539137	20	0.17257
47	0.12566	56	0.0444589	52	0.170249

56	0.12518	47	0.0330151	89	0.139157
26	0.10880	26	0.0285314	27	0.128016
30	0.09098	65	0.0281573	116	0.116835
9	0.08149	133	0.0234623	26	0.115961
33	0.08024	132	0.022598	34	0.112413
32	0.07959	21	0.0223288	5	0.107358
22	0.07903	107	0.0217067	56	0.104421
52	0.07777	27	0.021147	65	0.103961
27	0.07510	53	0.0204812	53	0.0994692
5	0.06298	121	0.019318	4	0.0987341
132	0.06001	52	0.0191321	93	0.0978349
4	0.05941	24	0.0186786	9	0.0972768
75	0.05711	28	0.0185364	30	0.096602
116	0.05366	33	0.0183399	118	0.0947606
65	0.05296	49	0.0183153	121	0.0935034
7	0.05254	116	0.0180679	22	0.0916174

В таблице 3.8 представлены результаты приближенного вычисления электрической центральности вершин графа с помощью метода последовательных приближений рекуррентной формулой (1.10) при $k = 2$, а также метода Монте-Карло при котором использовались в качестве исходных только 10% вершин. Для сравнения приведены результаты первых 20 вершин.

Вычисление электрической центральности методом последовательных приближений за две итерации выявило 6 первых вершин, но из первых двадцати вершин выявлено 12 вершин. Метод Монте-Карло с использованием 15 вершин ($\approx 10\%$) в качестве исходных выявил 15 вершин из 20 первых по значению электрической центральности при $\delta = 1$.

3.3. Ранжирование сайтов научных организаций

В качестве примера рассмотрим фрагмент веб-графа научных учреждений Российской академии наук (далее – веб-граф РАН) в их дореформенной версии, построенного с использованием базы данных внешних гиперссылок [2].

Веб-граф РАН представляет собой ориентированный граф с кратными ду-

гами без петель. По данным на март 2014 года он содержал 956 вершин, соответствующих сайтам научных отделений, центров, институтов, библиотек и др., связанных почти 39000 дугами, соответствующими гиперссылкам, связывающим эти сайты.

Неориентированный граф учреждений РАН представлен на рисунке 3.7. Данный неориентированный граф получен для проведения вычислительных экспериментов путем следующих действий:

1. оставлены только те вершины веб-графа, любая пара которых связана «встречными» дугами любой кратности;
2. все кратные встречные дуги заменены на ребра, вес которых равен меньшей из двух кратностей встречных дуг;
3. в полученном после первых двух шагов графе удалены все вершины (и соответствующие им ребра) не вошедшие в максимальную компоненту связности.

Также из построенного вышеописанным способом графа была удалена вершина, соответствующая сайту РАН с доменным именем *www.ras.ru*. Имея приблизительно 500 исходящих гиперссылок и 1300 входящих, сайт РАН занимает безусловное доминирующее положение в системе веб-сайтов учреждений РАН. Удаление вершины, соответствующей сайту РАН с доменным именем *www.ras.ru* она значительно упрощает сложность графа. Что позволяет нам более наглядно продемонстрировать более высокую чувствительность метода ранжирования с помощью электрической центральности по отношению к вершинам с определенными особенностями.

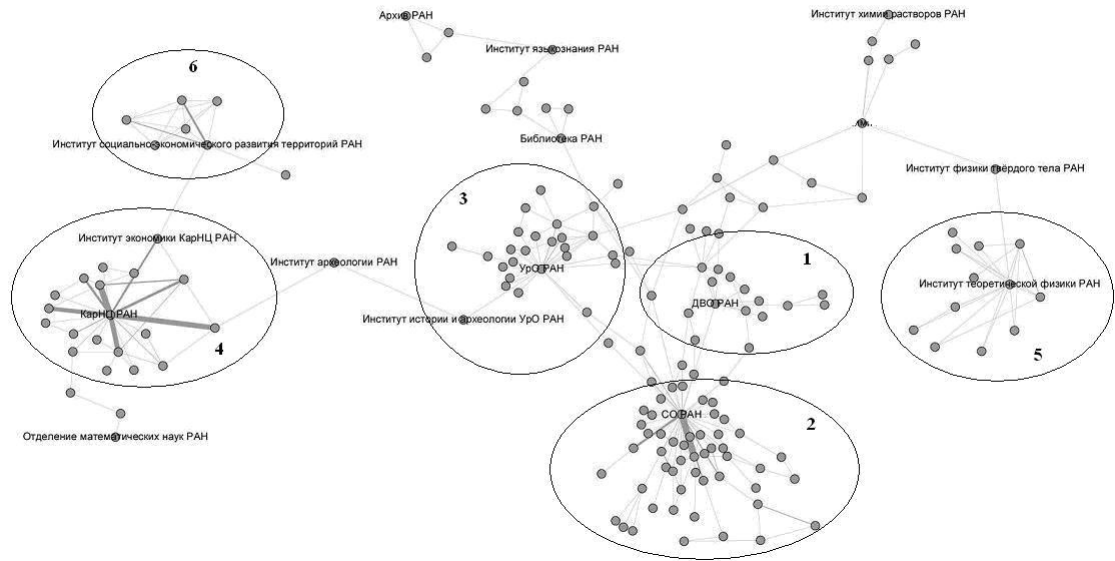


Рисунок 3.7. Неориентированный граф учреждений РАН.

Полученный таким образом граф содержит 169 вершин, связанных 279 ребрами. Вес ребер варьируется от 1 до 38, средний вес ребра равен 2.04.

На рисунке 3.7 некоторые вершины графа помечены названиями научных учреждений. Пронумерованными эллипсами выделены различные группы веб-сайтов такие как:

1-3. сайты научных учреждений, входящих в состав Дальневосточного, Сибирского и Уральского региональных отделений РАН (ДВО РАН, СО РАН и УрО РАН);

4. сайты научных учреждений, входящих в состав КарНЦ РАН;

5-6. сайты институтов (Институт социально-экономического развития территорий РАН, Институт теоретической физики РАН).

Ясно, что сайты, являющиеся «головными» в этих группах, должны иметь высокие значения PageRank (PR), представленной в работе [34]. Это подтверждается данными расчетов значения PageRank табл. 3.8 в которые попали все сайты учреждений, указанных в пп. 1-5.

Для этих же сайтов были сделаны расчеты значений электрической центральности и центральности по вектору Майерсона [61]. Сравнивая значения PR с ранжированием по значениям электрической центральности для $\delta = 0.3$

можно отметить, что 5 из 6 указанных головных сайтов попадают в первую десятку и в этом случае, а сайт www.vscs.ac.ru находится на 15-м месте.

В результате ранжирования вершин графа сайтов значением центральности по вектору Майерсона при $r = 0.9$ только 3 головных сайта попадают в первую десятку.

Таблица 3.9

Десять первых сайтов с наибольшим значением PR.

Название	URL	PR
Сибирское отделение РАН	www.sbras.nsc.ru	0.088
Карельский научный центр РАН	www.krc.karelia.ru	0.048
Институт вычислительных технологий СО РАН	www.ict.nsc.ru	0.038
Уральское отделение РАН	www.uran.ru	0.037
Институт теоретической физики им. Л.Д. Ландау РАН	itp.ac.ru	0.022
Дальневосточный геологический институт ДВО РАН	www.fegi.ru	0.017
Дальневосточное отделение РАН	www.febras.ru	0.013
Институт социально-экономического развития территорий РАН	www.vscs.ac.ru	0.013
Институт проблем химической физики РАН	www.icp.ac.ru	0.012
Институт прикладных математических исследований КарНЦ РАН	mathem.krc.karelia.ru	0.012

На рисунке 3.7 несложно выявить сайты, удаление которых ведет к образованию достаточно крупных несвязных компонент. В первую очередь, к ним относятся сайты Института археологии РАН, Института истории и археологии УрО РАН, Института экономики КарНЦ РАН, Библиотеки РАН и Института

физики твердого тела РАН. Значения рангов сайтов этих учреждений по PR, электрической центральности ($\delta = 0.3$) и вектору Майерсона ($r = 0.9$) приводятся в табл. 3.10.

Таблица 3.10

Сравнение рангов 5 учреждений РАН.

Название (URL)	Ранг по PR	Ранг по электрической центральности	Ранг по вектору Майерсона
Институт археологии РАН (www.archaeolog.ru)	76	41	3
Институт истории и археологии УрО РАН (www.ihist.uran.ru)	133	38	5
Институт экономики КарНЦ РАН (economy.krc.karelia.ru)	80	17	7
Библиотека РАН (www.rasl.ru)	151	19	35
Институт физики твердого тела РАН (www.issp.ac.ru)	137	23	22

Очевидно, что обнаружение указанных сайтов, имеющих важное значение в анализируемом графе РАН, более точно проводится с помощью методов, основанных на использовании понятий электрической центральности и вектора Майерсона, нежели с помощью PR. Например, это заметно на примере сайта Института физики твердого тела РАН, улучшающего свои позиции на 114 и 115 пунктов соответственно. На рис. 3.7 видно, что удаление этого сайта приведет к образованию двух несвязных компонент, к одной из которых относится группа,

состоящая из 12 сайтов Института теоретической физики РАН.

3.4. Онтологическая модель

На сегодняшний день общее число компетенций, практически в любой области деятельности человека, может быть настолько велико, что овладение всем множеством компетенций не представляется возможным. С другой стороны, благодаря общим базовым знаниям и компетенциям специалист может обладать высоким уровнем профессионализма. Это связано с тем, что компетенции могут быть взаимодополняющими и в некотором смысле взаимозаменяемыми.

Указанное свойство компетенций нашло свое отражением во множестве предлагаемых способов описания моделей компетенций. Например, в [37] модель компетенций описана методами нечетких множеств и для ранжирования кандидатов используется расстояние по Хеммингу между параметрами кандидата и идеального кандидата. В [46, 75] описаны модели компетенций, где учитываются уровни владения компетенцией. И соответственно при формировании команды учитываются уровни компетенций необходимые для решения той или иной задачи.

Совокупность знаний о некой конкретной предметной области можно представить в виде древовидной структуры, где вершины представляют основные понятия и термины рассматриваемой предметной области, а связи между вершинами - это иерархические отношения „целое-часть“. Данные иерархические структуры терминов и понятий, принятых в предметной области можно интерпретировать как онтологию [8]. Например, можно получить древовидную структуру (онтологию) путем декомпозиции понятий предметной области до некоторого уровня детализации в соответствии с решаемой задачей. Пример построения подобной онтологической модели представлен в [1].

В [9] модель компетенций представлена в виде онтологии, которая описывает взаимосвязь между компетенциями в определенной предметной области.

Кратко представим математическую модель.

Пусть $N = \{1, \dots, n\}$ - множество кандидатов, множество компетенций обозначим как $C = \{c_1, \dots, c_m\}$. Множество компетенции кандидата i обозначим как S_i , где $S_i \in C$, $i = 1, \dots, n$. Множество задач для решения, которых формируется команда, обозначим как $Z = \{Z_1, \dots, Z_t\}$, где $Z_j \in C$, $j = 1, \dots, t$.

Модель компетенций представлена онтологией O , которая определена деревом T , где корнем дерева T является предметная область, вершины дерева - это компетенции $C = \{C_1, \dots, C_m\}$, а ребра дерева указывают на иерархическую связь между понятиями-компетенциями разных уровней. Таким образом, онтологическую модель компетенций можно представить как $O = \langle T, C \rangle$.

Введем обозначения: L - высота онтологии O (высота дерева T), $l(c_i)$ - длина пути от корня до компетенции c_i .

Для более удобного формального описания онтологии O закодируем все компетенции в ней. Для этого пронумеруем все вершины-потомки онтологии O , начиная с 1 относительно родительской вершины. Таким образом, мы однозначно закодируем все компетенции с сохранением информации об отношениях между ними.

Определение 3.1. Адресом компетенции c_i называется вектор $(a_1, a_2, \dots, a_l, a_{l+1}, \dots, a_L)$ длины L , где элементы a_1, a_2, \dots, a_l равны номерам вершин, лежащим на пути от корня к компетенции c_i , если $l(c_i) < L$, то элементы $a_{l(c_i)+1}, \dots, a_L$ равны 0.

Тогда расстояние $\rho(c_i, c_j)$ между двумя компетенциями $c_i = (a_1, a_2, \dots, a_L)$ и $c_j = (b_1, b_2, \dots, b_L)$ в онтологии O можно определить следующим образом:

$$\begin{aligned} \rho(c_i, c_j) = & \sum_{i=1}^L I\{a_i \neq b_i \cap (a_i = 0 \cup b_i = 0)\} + \\ & + 2 \sum_{i=1, j < i}^L ((L - i) I\{a_i \neq b_i \cap (a_j = b_j \cup i = 1) \cap (a_i \neq 0 \cap b_i \neq 0)\}) \end{aligned} \quad (3.1)$$

Чем больше расстояние между компетенциями c_i и c_j в онтологии O , тем больше разница между ними. Компетенция c_i является более ценной, если она

находится ближе ко всем остальным компетенциям в онтологии O , т.е. среднее расстояние до всех остальных компетенций $c_j \neq c_i$. Здесь можно использовать меру близости (closeness centrality) [21] для ранжирования всех компетенций в онтологии O .

Для ясности и отражения сути меры близости для нашей онтологической модели компетенций введем следующее определение.

Определение 3.2. Мерой покрытия онтологии O компетенцией c_i называется величина:

$$\gamma(c_i) = \frac{1}{\sum_{c_j \neq c_i, c_j \in C} \rho(c_i, c_j)} \quad (3.2)$$

где $\rho(c_i, c_j)$ - расстояние между компетенциями $c_i, c_j \in C$.

Используя меру покрытия, можно ранжировать все компетенции в онтологии.

Определение 3.3. В онтологии O расстоянием между набором компетенций $S_i \subset C$ и компетенцией $c' \in C$ называется величина $\rho(S_i, c') = \min_{c'' \in S_i} \rho(c'', c')$.

Данное определение расстояния можно рассматривать как степень владения компетенцией c' кандидата i . Если $\rho(S_i, c') = 0$, то следует, что $c' \in S_i$, т.е. кандидат i владеет компетенцией c' . Допустим у кандидата i $\rho(S_i, c') = 2$ и $\rho(S_i, c'') = 4$, тогда можно предложить следующую интерпретацию: кандидат i не владеет компетенциями c' и c'' , но для овладения компетенцией c' ему необходимо время равное 2 и соответственно, для овладения c'' время равное 4.

Определение 3.4. Мерой покрытия онтологии O набором компетенций S_i называется величина:

$$\gamma(S_i) = \frac{1}{\sum_{c' \notin S_i, c' \in C} \rho(S_i, c')} \quad (3.3)$$

где $\rho(S_i, c')$ - расстояние между набором компетенций $S_i \subset C$ и компетенцией $c' \in C \setminus S_i$.

Функция γ является неаддитивной.

Мера покрытия онтологии O определенная для набора компетенций позволяет, например, оценить общий профессиональный уровень кандидата в опре-

деленной области. Чем выше мера покрытия, тем лучше кандидат. Мера покрытия соответствует решению задачи выбора наилучшего кандидата, когда имеется такое множество целевых задач, для решения которых в равной степени может понадобиться каждая компетенция из O .

Предположим, что задано некоторое множество задач Z , которые должен выполнять работник. Требуется найти наилучшего подходящего работника из предлагаемых кандидатов.

Пусть множество задач Z определено множеством компетенций необходимых для успешного решения задач из Z , т.е. $Z = \{Z_1, \dots, Z_t\} = \bigcup_{k=1}^t C_k = \{c_1^z, \dots, c_r^z\}$, где $C_k \subseteq C$ - множество компетенций необходимых для решения задачи Z_k .

Введем характеристическую функцию:

$$v_i(x) = \sum_{\rho=0}^{2L} \alpha_\rho x^\rho, \quad (3.4)$$

где α_ρ - число компетенций из $\{c_1^z, \dots, c_r^z\}$ на расстоянии ρ от ближайшей компетенции S_i кандидата i в онтологии O , $x \in [0, 1]$.

Отметим, что $\max_{c_i, c_j \in C} \rho \leq 2L$, т.к. наибольшее расстояние между двумя компетенциями не превышает $2L$.

Рассмотрим многокритериальную оценку кандидата.

Пусть все компетенции множества $Z = \{c_1^z, \dots, c_r^z\}$ имеют разные веса w_k^z , где w_k^z - вес компетенции c_k^z . Тогда получим следующую оценку кандидата i :

$$E_i = \frac{1}{r} \sum_{k=1}^r w_k^z \rho(S_i, c_k) \quad (3.5)$$

Далее представим формализованную постановку задачи формирования команды.

При заданных значениях $\langle Z, N, O, q \rangle$, где q - число членов в команде, требуется определить оптимальную команду из множества кандидатов N .

Предположим, что команда владеет всеми компетенциями ее членов, т.е. множество компетенций команды состоит из объединения множеств компетен-

ций ее членов. Таким образом, приведенные критерии оценки отдельных кандидатов (3.2)-(3.4) справедливы и для оценки оптимальности формируемой команды.

Пример 3.1.

На рисунке 3.8 представлен пример упрощенной онтологии компетенций из предметной области: «Олимпиада по программированию». Для решения конкретной задачи выдаваемой на олимпиаде команда должна владеть определенным набором компетенций необходимых для ее решения. Если команда не владеет какой-нибудь компетенцией, то она может компенсировать этот недостаток другими компетенциями близкими к недостающей и все же решить поставленную задачу. Далее на примерах будем исследовать данную онтологию компетенций.

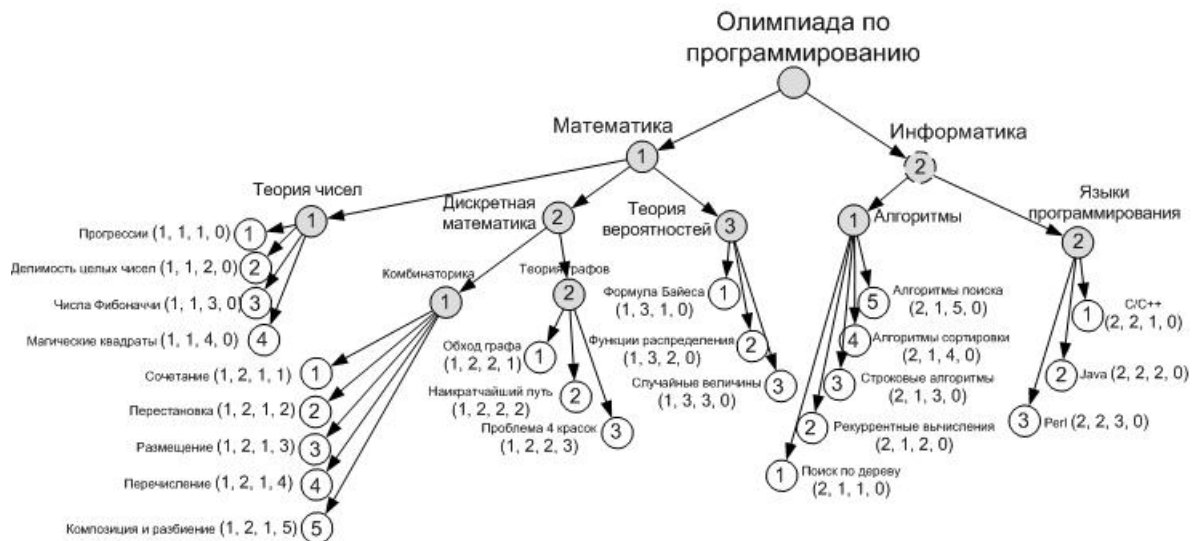


Рисунок 3.8. Пример онтологии «Олимпиада по программированию».

Рассмотрим задачу ранжирования, где ценность команды определяется относительно всей онтологии компетенции. Или другими словами в случае, когда множество целевых задач задано так, что для их решения в равной степени может понадобиться каждая компетенция из O .

Для этого используем меру покрытия, где оптимальной командой будет та, которая обеспечивает максимальное покрытие онтологии.

Пусть каждый кандидат владеет только одной компетенцией: $S_1 = (1, 1, 1, 0)$, $S_2 = (1, 2, 1, 1)$, $S_3 = (1, 2, 2, 1)$, $S_4 = (1, 3, 2, 0)$, $S_5 = (2, 1, 5, 0)$, $S_6 =$

$(2, 2, 2, 0)$ и $S_7 = (2, 1, 1, 0)$.

Результаты вычисления мер покрытия по (3.2) кандидатов представлены в таблице 3.11.

Таблица 3.11

Меры покрытия кандидатов.

Кандидат	S_1	S_2	S_3	S_4	S_5	S_6	S_7
Мера покрытия (γ)	$\frac{1}{106}$	$\frac{1}{111}$	$\frac{1}{115}$	$\frac{1}{108}$	$\frac{1}{118}$	$\frac{1}{122}$	$\frac{1}{118}$

Вычислим электрическую меру близости с помощью выражения (1.16). Корректное применение электрической меры близости в контексте данной онтологической модели, предполагает ряд уточнений:

1. электрическая мера близости определена только для компетенций, т.е. для листьев дерева онтологии;
2. при расчете электрической меры близости учитывается только разница абсолютных потенциалов между листьями, т.е. между компетенциями.

Таблица 3.12

Электрические меры близости кандидатов.

Кандидат	S_1	S_2	S_3	S_4	S_5	S_6	S_7
Мера близости (CF_C) при $\delta = 1$	2.6147	2.6567	2.5562	2.5595	2.6552	2.5541	2.6552

Результаты вычислений электрической меры близости при $\delta = 1$ по формуле (3.2) кандидатов представлены в таблице 3.12. По результатам ранжирования компетенций в онтологии мерой близости по (3.2) и мерой покрытия можно сделать следующие выводы. Результаты ранжирования существенно отличаются, несмотря на то, что в обоих случаях компетенция S_6 получила наименьший ранг, а компетенции S_1 и S_2 получили высокие ранги. В результате ранжирования электрической меры близости при $\delta = 1$, компетенция S_2 получила ранг

выше, чем компетенция $S2$. Более того компетенции $S5$ и $S7$ получили ранг выше $S1$, за счет близости к корню и большего числа компетенций в группе „Алгоритмы“.

Таким образом, в онтологической модели компетенций можно успешно использовать электрическую меру близости. К тому же есть возможность улучшить адекватность модели за счет подбора значения δ в каждом конкретном случае.

Заключение

В диссертации предлагается ряд математических моделей ранжирования вершин в графах коммуникационных сетей, исследуются свойства данных моделей.

В работе представлен новый способ определения центральности для взвешенных графов на основе правил Кирхгофа, которые представляют собой систему линейных уравнений. Предлагаемый способ сохранил полезные свойства мер центральностей основанных на модели электрических цепей, но имеет сравнительно невысокую вычислительную сложность равную сложности вычисления обратной матрицы, т.е. $O(n^3)$.

Для некоторых частных видов графов (полный двудольный граф и частный случай взвешенной звезды) электрическая центральность вершин найдена в аналитическом виде.

В работе представлен алгоритм приближенного вычисления предлагаемой электрической центральности методом последовательных приближений для ранжирования вершин в графах большой размерности. Представлена возможность применения стохастического метода Монте-Карло для поиска электрической центральности.

На ряде примеров сделано сравнение предлагаемого способа определения центральности с традиционным методом ранжирования PageRank, где показаны некоторые преимущества электрической центральности.

Построена модель с использованием методов кооперативной теории игр, где характеристическая функция строиться на основе значений электрической центральности вершин. С помощью программной реализации на ЭВМ проведены численные эксперименты.

Представлены результаты применения предложенного нами способа определения электрической центральности для анализа значимости вершин на примерах графа сообщества „Теория игр“ в социальной сети ВКонтакте, графа

публикаций Math-Net.ru написанных в соавторстве, а также на ряде графов транспортных сетей. Кроме того, представлен пример ранжирования сайтов научных учреждений Российской академии наук.

Разработаны модели и комплекс программ в пакете Mathematica для анализа коммуникационных сетей.

Полученные результаты носят как теоретический, так и прикладной характер.

Литература

1. Вдовицын В.Т., Лебедев В.А. Онтологическое моделирование контента электронной библиотеки КарНЦ РАН // Труды КарНЦ РАН. 2010. №3. Серия «Математическое моделирование и информационные технологии». Вып. № 1. С. 11–19.
2. Головин А.С., Печников А.А. База данных внешних гиперссылок для исследования фрагментов Веба // Информационная среда вуза XXI века: материалы VII Всероссийской научно-практической конференции (23-27 сентября 2013 г.). 2013. С. 55-57.
3. Жижченко А.Б., Изаак А.Д. Информационная система Math-Net.Ru. Применение современных технологий в научной работе математика // Успехи Математических Наук. 2007. Т. 62, № 5(377). С. 107–132.
4. Жижченко А.Б., Мазалов В.В., Цынгугев Б.Т. Ранжирование вершин графа публикаций математического портала Math-Net.Ru // Труды Карельского научного центра РАН. 2015. №5. С. 34–41.
5. Мазалов В.В., Цынгугев Б.Т. Программа вычисления центральности вершин графа на основе правил Кирхгофа. 2015. Свидетельство о государственной регистрации программы для ЭВМ № 2015617367 от 08.07.2015.
6. Мазалов В.В., Цынгугев Б.Т. Ранжирование вершин взвешенных графов коммуникационных сетей на основе правил Кирхгофа // Расширенные тезисы представленные на международном семинаре „Сетевые игры и менеджмент“. Петрозаводск, Россия, июль 5-7. 2015. С. 35-37.
7. Трухина Л.И. Кооперативные решения в задачах анализа информационных сетей: Дис. ... канд. физ.-мат. наук: 05.13.18 - Математическое моделирование, численные методы и комплексы программ / Забайкальский государственный гуманитарно-педагогический университет им. Н.Г. Чернышевского. Чита, 2013. 110 с.

8. Фазлиев А.З. Рассуждения о понятии „онтология“ // Онтологическое моделирование. Труды симпозиума. Звенигород. 2008. С. 278-296.
9. Цынгугев Б.Т. Задача формирования команды на основе онтологической модели компетенций // Учёные записки Забайкальского государственного университета серия Физика, математика, техника, технология. 2014. №3(56) С. 110-116.
10. Цынгугев Б.Т. Многошаговая игра формирования сети с экспоненциальной выплатой // Расширенные тезисы представленные на международном семинаре „Сетевые игры и менеджмент“. Петрозаводск, Россия, июнь 23-25. 2013. С. 98-101.
11. Цынгугев Б.Т. Новый способ вычисления центральности для взвешенных графов коммуникационных сетей на основе правил Кирхгофа // Сборник тезисов представленные на международной конференции „Дифференциальные уравнения и математическое моделирование“. Улан-Удэ, Байкал, Россия, июнь 22-27. 2015. С. 316-317.
12. Цынгугев Б.Т. Многошаговая игра формирования сети с экспоненциальной выплатой // Математический анализ и его приложения. 2013. №11. С. 59-71.
13. Aumann R., R. Myerson. Endogenous formation of links between players and coalitions: an application of the Shapley value // In: The Shapley value, Cambridge University Press. 1988. Pp. 175–191.
14. Avrachenkov K., Litvak N., Medyanikov V., Sokol M. Alpha current flow betweenness centrality // In Proceedings of WAW 2013, also LNCS v.8305. 2013. Pp. 106-117.
15. Avrachenkov K., Litvak N., Nemirovsky D., Smirnova E., Sokol M. Quick detection of top-k personalized pagerank list // In Proceedings of WAW 2011, LNCS v.6732.2011. Pp. 50-61.
16. Avrachenkov K., Filar J., Howlett P. Analytic Perturbation Theory and its Applications // SIAM. 2013.

17. Avrachenkov K., Litvak N., Nemirowsky D., Osipova N. Monte carlo methods in PageRank computation: When one iteration is sufficient // SIAM J. Numer. Anal. Vol. 45(2). 2007. Pp.890–904.
18. Avrachenkov K.E., Mazalov V.V., Tsynguev B.T. Beta Current Flow Centrality for Weighted Networks // Lecture Notes in Computer Science, vol. 9197. Computational Social Networks, Springer. 2015. Pp. 216-227.
19. Banzhaf J.F. Weighted voting doesn't work: mathematical analysis // Rutgers Law Review. 1976. №19. Pp 317-343.
20. Bastian M., Heymann S., Jacomy M. Gephi: an open source software for exploring and manipulating networks // International AAAI Conference on Weblogs and Social Media. 2009.
21. Beauchamp M.A. An improved index of centrality // Behavioral Science. 1965. Vol. 10. P.161-163.
22. Bianchini M., Gori M., Scarselli F. Inside PageRank //ACM Trans. on Internet Technology. Vol. 5. 2005. Pp. 92–128.
23. Borgatti S.P., Everett M.G. and Freeman L.C. Ucinet for Windows: Software for Social Network Analysis // MA: Analytic Technologies. Harvard. 2002.
24. Borm P., Owen G., Tijs S. On the position value for communication situations // SIAM J. on Discrete Math. Vol. 5(3). 1992. Pp. 305-320.
25. Borm P., van den Nouweland A., Tijs S. Cooperation and communication restrictions: a survey // In: Imperfections and Behavior in Economic Organizations. Kluwer Acad. Publ. Boston. 1994.
26. Boccaletti S., Latora V., Moreno Y., Chavez M., Hwang D.U. Complex networks: Structure and dynamics // Physics Reports-Review Section of Physics Letters. Vol. 424. 2006. Pp. 175–308.
27. Borgatti S.P. Centrality and Network Flow // Social Networks (Elsevier). Vol. 27. 2005. Pp. 55–71.

28. Brandes U. A faster algorithm for betweenness centrality // Journal of Mathematical Sociology. Vol. 25. 2001. Pp. 163-177.
29. Brandes U., Fleischer D. Centrality measures based on current flow // In Proceedings of the 22nd annual conference on Theoretical Aspects of Computer Science. 2005. Pp. 533-544.
30. Breiger R. The duality of persons and groups // Social Forces. Vol. 53. 1974. Pp. 181-190.
31. Breiger R., Pattison P. Cumulated social roles: The duality of persons and their algebras // Social Networks. Vol. 8. 1986. Pp. 215-256.
32. Breyer L.A. Markovian page ranking distributions: some theory and simulations // Technical report. 2002.
33. Breyer L.A., Roberts G.O. Catalytic perfect simulation // Methodol. Comput. in Appl. Probab., Vol. 3. 2001. Pp. 161-177.
34. Brin S., Page L. The anatomy of a large-scale hypertextual Web search engine. Computer Networks and ISDN Systems. 1998. Vol. 30(17). P. 107-117.
35. Brin S., Page L., Motwami, Winograd T. The PageRank citation ranking: bringing order to the Web. Stanford University Technical Report. 1998.
36. Calvo E., Lasaga J., van den Nouweland A. Values of games with probabilistic graphs // Math. Social Sci., Vol. 37. 1999. Pp. 79-95.
37. Canós L., Casasús T., Crespo E., Lara T. and Pérez J.C. Personnel selection based on fuzzy methods // Revista De Matemática: Teoría Y Aplicaciones. 2011. Vol. 18, No. 1. Pp. 177-192.
38. Chebukov D., Izaak A., Misurina O., Pupyrev Yu., Zhizhchenko A. Math-Net.Ru as a digital archive of the Russian mathematical knowledge from the XIX century to today // Lecture Notes in Computer Science. 2013. Vol. 7961. P. 344-348.
39. Davis A., a et al. Deep South // University of Chicago Press. 1941.

40. Deegan J., Packel E.W. A new index of power for simple n-person games // International Journal of Game Theory. 1978. № 7. Pp. 113-123.
41. Fogaras D., Racz B. Towards scaling fully personalized PageRank // In Proceedings of the 3-rd International Workshop on Algorithms and Models for the Web-Graph, WAW. 2004.
42. Freeman L.C. A set of measures of centrality based on betweenness // Sociometry. Vol. 40. 1977. Pp. 35-41.
43. Freeman L.C. Centrality in Social Networks Conceptual Clarification // Social Networks. Vol. 1. 1979. Pp. 215-239.
44. Freeman L.C., Borgatti S.P., White D.R. Centrality in valued graphs: A measure of betweenness based on network flow // Social Networks. Vol. 13. 1991. Pp. 141-154.
45. Goh K.-I., Oh E., Kahng B., Kim D. Betweenness centrality correlation in social networks // Phys. Rev. E. Vol. 67. 2003. 017101.
46. Hlaotittinun O., Bonjour E., Dulmet M. A team building approach for competency development // Proceedings of the IEEE International Conference on Industrial Engineering and Engineering Management. 2007. P. 1004-1008.
47. Holler M.J. Forming coalitions and measuring voting power // Political Studies. 1982. №30. Pp. 262-271.
48. Ipsen I.C.F., Kirkland S. Convergence Analysis of an improved PageRank algorithm // North Carolina State University Technical Report CRSC-TR04-02. 2004.
49. Jackson M.O. Allocation rules for network games // Games and Econ. Behav. Vol. 51(1). 2005. Pp. 128-154.
50. Jackson M.O., Wolinsky J. A strategic model of social and economic networks // J. Econ. Theory. Vol. 71(1). 1996. Pp. 44-74.
51. Jackson M.O. Social and economic networks // Princeton University Press. 2008.

52. Jamison R.E. Alternating Whitney sums and matchings in trees // part 1. Discrete Math. Vol. 67. 1987. Pp. 177-189.
53. Jeh G., Widom J. Scaling personalized web search // In Proceedings of the 12th World Wide Web Conference. 2003.
54. Kamvar S.D., Haveliwala T.H., Golub G.H. Adaptive methods for the computation of PageRank // Lin. Alg. Appl. Vol. 386. 2004. Pp.51-65.
55. Kamvar S.D., Haveliwala T.H., Manning C.D., Golub G.H. Extrapolation Methods for Accelerating PageRank Computations // In Proceedings of the 12-th International World Wide Web Conference. 2003.
56. Kent D. The rise of the Medici: Faction in Florence // Oxford University Press. 1978. Pp. 1426-1434.
57. Korzun D., Gurtov A. Structured Peer-to-Peer Systems: Fundamentals of Hierarchical Organization, Routing, Scaling, and Security // Springer. 2013.
58. Langville A.N., Meyer C.D. Deeper Inside PageRank // Internet Mathematics. Vol. 1. No.3. 2004. Pp.335-400.
59. Langville A.N., Meyer C.D. Updating PageRank with Iterative Aggregation // In Proceedings of the 13-th World Wide Web Conference. 2004.
60. Mazalov V. Mathematical Game Theory and Applications // Wiley. 2014.
61. Mazalov V.V., Trukhina L.I. Generating functions and the Myerson vector in communication networks // Discrete Mathematics and Applications. Vol. 24(5). 2014. Pp. 295-303.
62. Myerson R.B. Graphs and cooperation in games // Math. Oper. Res. Vol 2. 1977. Pp. 225-229.
63. Nakao K. Distribution of measures of centrality: Enumerated distributions of Freeman's graph centrality measures // Connections. № 13(3). 1990. Pp. 10-22.
64. Newman M. E. J. The structure and function of complex networks // SIAM Review. Vol 45. 2003. Pp. 167-256.

65. Newman M.E.J. A measure of betweenness centrality based on random walks // Social networks. Vol. 27. 2005. Pp. 39-54.
66. Newman M.E.J. Networks: An Introduction // Oxford University Press, Oxford, UK. 2010.
67. Opsahl T., Agneessens F., Skvoretz J. Node centrality in weighted networks: generalizing degree and shortest paths // Social Networks. Vol. 32. 2010. Pp. 245-251.
68. Padgett J.F. and Ansell C.K. Robust action and the rise of Medici // Am. J. Sociol. Vol. 98. 1993. Pp 1259-1319.
69. Slikker M. Link monotonic allocation schemes // Int. Game Theory Review Vol. 7(4). 2005. Pp. 473-489.
70. Slikker M., Gilles R.P., Norde H., Tijs S. Directed networks, allocation properties and hierarchy formation// Math. Social Sci. Vol. 49(1). 2005. Pp. 55-80.
71. Shapley L.S. A value for n -person games // Annals of Mathematical studies. 1953. №28. Pp. 307-317.
72. Shapley L.S., Shubik M. A method for evaluating the distribution of power in a committee system // American Political Science Review. 1954. №48. Pp. 787-792.
73. Strogatz S.H. Exploring Complex Networks // Nature. Vol. 410 (6825).2001. Pp. 268–276.
74. Talman D., Yamamoto Y. Average tree solutions and subcore for acyclic graph games // J. Oper. Res. Soc. Japan, Vol. 51(3). 2008. Pp. 187-201.
75. Tarasov V. Ontology-based Approach to Competence Profile Management // Journal of Universal Computer Science. 2012. Vol. 18. No. 20. Pp. 2893-2919.
76. Wasserman S., Faust K. Social Network Analysis: Methods and Applications // Cambridge University Press. 1994.

- 77. Wilhelm T., Kim J. What is a complex graph? // Physica A. Vol. 387. 2008. Pp. 2637–2652.
- 78. Yamamoto A., Asahara D., Itao T., Tanaka S., Suda T. Distributed pagerank: A distributed reputation model for open peer-to-peer networks // In: Proc. 2004 Int'l Symposium on Applications and the Internet Workshops, IEEE Computer Society. 2004. Pp. 389-394.
- 79. Zachary W.W. An information flow model for conflict and fission in small groups // Journal of Anthropological Research. Vol. 33. 1977. Pp. 452-473.

Приложение А

Листинги программного комплекса

1. Программный комплекс, разработанный в пакете Mathematica

Данный программный комплекс позволяет вычислить меру центральности и меру близости в модели электрических цепей, а также вычислить их приближенные значения методом последовательных приближений.

```

MatrixAdjacency=Import["Path to file of adjacency matrix","Table"];
MatrixN=Length[MatrixAdjacency];
g=GraphPlot[ MatrixAdjacency]
B=IdentityMatrix[MatrixN] ;
MatrixG[M_,N_]:=
Block[{tempMatr=Table[0,{N},{N}],i,j},
For[i=1,i<=N,i++,
For[j=1,j<=N,j++,
tempMatr[[i,i]]+=M[[i,j]] ]];
tempMatr=tempMatr-M;
tempMatr]

DifferenceOfPotentials[absolutePotentials_,M_,N_]:=
Block[{tempDifferencePotentials=Table[0,{N}],i,j},
For[i=1,i<=N,i++,
For[j=1,j<=N,j++,
tempDifferencePotentials[[i]]+=
M[[i,j]]*Abs[Part[absolutePotentials,i]-Part[absolutePotentials,j]]]];
tempDifferencePotentials]

CentralityOfCurrentFlow[M_,N_,delta_]:=
Block[{tempPotentials=Table[0,{N}],s,
D1=Inverse[MatrixG[M,N]+delta*B]},
For[s=1,s<=N,s++,

```

```

tempPotentials+=0.5*DifferenceOfPotentials[D1.B[[s]],M,N]];
(tempPotentials+1/2)/N]
  ClosenessOfCurrentFlow[M_,N_,delta_]:=
Block[{s, j, D1 = Inverse[MatrixG[M, N] + delta*B],
potentials = Table[0, {N}]}],
For[s = 1, s <= N, s++,
For[j = 1, j <= N, j++,
potentials[[s]] += D1[[s, s]] - D1[[s, j]]
]]; (N - 1)/potentials]
  MatrixD[M_,N_]:=
Block[{tempMatr=Table[0,{N},{N}],i,j},
For[i=1,i<=N,i++,
For[j=1,j<=N,j++,
tempMatr[[i,i]]+=M[[i,j]] ]];
tempMatr]
  VectorPi[M_,N_,n1_]:=
Block[{tempMatr=Inverse[MatrixD[M,N]].M},
tempMatr=N[MatrixPower[tempMatr,n1]];
N[tempMatr]]
  D1[M_,N_,delta_]:=
Block[{d=MatrixD[M,N]},
Inverse[d+delta*B]]
  D2[M_,N_,delta_]:=
Block[{d=MatrixD[M,N]},
Inverse[d+delta*B].d]
  Phi[M_,N_,delta_,s_]:=
Block[{k,sum=0},
For[k=0,k<=N,k++,
sum+=MatrixPower[D2[M,N,delta].Inverse[MatrixD[M,

```



```

N]].M,k].(D1[M,N,delta].B[[s]]);
sum]

CentralityPiApprox[M_, N_, delta_] :=
Block[{tempPotentials=Table[0,{N}],
s,i1,VPi=VectorPi[60],d=MatrixD[M,N],sumi=0},
For[i1=1,i1<=N,i1++,
sumi=SumD[i1,N];
For[s=1,s<=N,s++,
tempPotentials[[i1]]+=1+VPi[[s,s]]*sumi/(delta*(d[[s,s]]+delta))]];
tempPotentials/(2*N)]

PhiOneIteration[M_,N_,X_,delta_] :=
D2[M,N,delta].Inverse[MatrixD[M,N]].M.X+D1[M,N,delta]

CentralityOfCurrentFlowIterations[M_,N_,delta_,IterN_] :=
Block[{PhiTemp=Nest[PhiOneIteration[M,N,#,delta]&,
Table[0,{N},{N}], IterN], tempPotentials=Table[0,{N}],s},
For[s=1,s<=N,s++,
tempPotentials+=0.5*DifferenceOfPotentials[PhiTemp.B[[s]],N]];
(tempPotentials+1/2)/N]

```

2. Программа, разработанная на языке программирования Java

Данная программа предназначена для построения графа публикаций по данным сайта Math-Net.ru. В качестве HTML-парсера использовалась библиотека jsoup-1.8.1.jar (<http://jsoup.org/news/release-1.8.1>). Представлен фрагмент кода программы.

```

public class DataMatrix {
private int countOfVertex=0;
public ArrayList<Vertex> v=null;
public Integer [[[ matrixAdjacency;

    public DataMatrix(){
v=new ArrayList<Vertex>();}

```

```

    public void add(String userid, String username){
Vertex v1=new Vertex(userid,username);
v1.id=countOfVertex++;
v.add(v1);}

    public int getCountOfVertex(){
this.matrixAdjacency=new Integer[countOfVertex][countOfVertex];
return countOfVertex;}

    public void getCountOfVertex(int countOfVertex){
this.matrixAdjacency=new Integer[countOfVertex][countOfVertex];
this.countOfVertex=countOfVertex; }

    public void printMatrixAdjacency(){
for(int i=0;i<this.countOfVertex;i++) {
for(int j=0;j<this.countOfVertex;j++)
System.out.print(this.matrixAdjacency[i][j]+"");
System.out.println();}}

    class Vertex{
String userid, username;
public int id;
public Vertex(String userid, String username){
this.userid=userid;
this.username=username; }} }

    public static void fillMatrix(ArrayList<String> authorsSet,
ArrayList<String> papers) throws
InterruptedException, FileNotFoundException, IOException{ int i,j;
int countOfAuthors=authorsSet.size();
Integer [][] matrixAdjacency = new Integer[countOfAuthors][countOfAuthors];
String urlPaper="http://www.mathnet.ru";
for(i=0;i<countOfAuthors;i++) {
for(j=0;j<countOfAuthors;j++)

```

```

matrixAdjacency[i][j]=0;}
ArrayList<String> tempListAuthors;
ArrayList<String> tempListCoauthors;
for (String paper : papers) {
tempListAuthors=getCoauthors(paper,urlPaper);
System.out.println(paper);
for (String coauthor : tempListAuthors) {
tempListCoauthors=new ArrayList(tempListAuthors);
tempListCoauthors.remove(coauthor);
for (String otherCoauthor : tempListCoauthors) {
i=authorsSet.indexOf(coauthor);
j=authorsSet.indexOf(otherCoauthor);
matrixAdjacency[i][j]++;
matrixAdjacency[j][i]++; }}
System.out.println(tempListAuthors.size());
Thread.sleep(2000); }
System.out.println("Матрица смежности графа соавторства:");
for(i=0;i<countOfAuthors;i++) {
for(j=0;j<countOfAuthors;j++)
System.out.print(matrixAdjacency[i][j]+"");
System.out.println(); }

    public static ArrayList<String> getPapers(String id, String urlPerson){
Document doc;
ArrayList<String> papers=new ArrayList();
try {
doc = Jsoup.connect(urlPerson+id).get();
Elements links = doc.select("a[href]");
for (Element link : links) {
if(link.attr("href").contains("/rus/")&&!link.attr("href").contains("/rus/person"))

```

```

{
papers.add(link.attr("href")); }}
} catch (IOException e) {
e.printStackTrace(); }
return papers;}

    public static ArrayList<String> getCoauthors(String idPaper,
String urlPaper){
Document doc;
ArrayList<String> authors=new ArrayList();
if(idPaper.startsWith("http://mi.mathnet.ru"))
idPaper=idPaper.split("http://mi.mathnet.ru")[1];
try {
doc = Jsoup.connect(urlPaper+idPaper).get();
Elements links = doc.select("a[href]");
for (Element link : links) {
if(link.attr("href").contains("personid")) {
authors.add(link.attr("href").split("personid=")[1]); }}
} catch (IOException e) {
e.printStackTrace();}
return authors; }

```

Приложение Б

Копия свидетельства о регистрации программы

РОССИЙСКАЯ ФЕДЕРАЦИЯ



СВИДЕТЕЛЬСТВО

о государственной регистрации программы для ЭВМ

№ 2015617367

**Программа вычисления центральности вершин графа на
основе правил Кирхгофа**

Правообладатель: *Федеральное государственное бюджетное
учреждение науки Институт прикладных математических
исследований Карельского научного центра Российской академии
наук (ИПМИ КарНЦ РАН) (RU)*

Авторы: *Мазалов Владимир Викторович (RU),
Цынгугев Булат Тимурович (RU)*



Заявка № 2015614244

Дата поступления 21 мая 2015 г.

Дата государственной регистрации

в Реестре программ для ЭВМ 08 июля 2015 г.

Врио руководителя Федеральной службы
по интеллектуальной собственности

Л.Л. Кирий