

ИССЛЕДОВАНИЕ РАНЖИРОВАНИЯ В ТЕМПОРАЛЬНЫХ СЕТЯХ**Киселев Михаил Альбертович¹, Прогулова Татьяна Борисовна²**¹Студент;

Государственный университет «Дубна»;

Россия, 141980, Московская обл., г. Дубна, ул. Университетская, 19;

e-mail: miha.kiseleff@yandex.ru.

²Кандидат технических наук, доцент;

Государственный университет «Дубна»;

Россия, 141980, Московская обл., г. Дубна, ул. Университетская, 19;

e-mail: progulova@yahoo.com.

Работа посвящена исследованию ранжирования во время-зависимых сетях. В работе проведен обзор типов и классов сетей, а также основных метрик центральностей, подходов к ранжированию и алгоритмов ранжирования. В рамках сетей Бианкони-Барабаси изучалась корреляция внутренней пригодности узлов, влияющей на эволюцию сети, и результатов ранжирования. В работе сформулированы выводы касательно взаимозаменяемости некоторых центральностей и целесообразности использования конкретных центральностей и подходов к ранжированию. Разработанное программное обеспечение может быть использовано для расширения и углубления исследований ранжирования в сложных сетях.

Ключевые слова: сложные сети, темпоральные сети, ранжирование, метрики центральности.

Для цитирования:

Киселев М. А., Прогулова Т. Б. Исследование ранжирования в темпоральных сетях // Системный анализ в науке и образовании: сетевое научное издание. 2025. № 3. С. 17-32. EDN: LFNYIM. URL: <https://sanse.ru/index.php/sanse/article/view/678>.

RESEARCH OF RANKING IN TEMPORAL NETWORKS**Kiselev Mikhail A.¹, Progulova Tatiana B.²**¹Student;

Dubna State University;

19 Universitetskaya Str., Dubna, Moscow region, 141980, Russia;

e-mail: miha.kiseleff@yandex.ru.

²PhD in Engineering sciences, associate professor;

Dubna State University;

19 Universitetskaya Str., Dubna, Moscow region, 141980, Russia;

e-mail: progulova@yahoo.com.

The paper is devoted to the study of ranking in time-dependent networks. The paper provides an overview of the types and classes of networks, as well as the main centrality metrics, ranking approaches, and ranking algorithms. Within the framework of Bianconi-Barabasi networks, the correlation of the intrinsic fitness of nodes, which affects the evolution of the network, and the ranking results was studied. The paper draws conclusions regarding the interchangeability of some centralities and the expediency of using specific centralities and ranking approaches. The developed software can be used to expand and deepen ranking research in complex networks.

Keywords: complex networks, temporal networks, ranking, centrality metrics.



Статья находится в открытом доступе и распространяется в соответствии с лицензией Creative Commons «Attribution» («Атрибуция») 4.0 Всемирная (CC BY 4.0) <https://creativecommons.org/licenses/by/4.0/deed.ru>

For citation:

Kiselev M. A., Progulova T. B. Research of ranking in temporal networks. *System analysis in science and education*, 2025;(3):17-32 (in Russ). EDN: LFNYIM. Available from: <https://sanse.ru/index.php/sanse/article/view/678>.

Введение

В условиях постоянно возрастающего объема информации все более актуальным становится ее ранжирование. Итоговые рейтинги обуславливают выбор товаров и их поставщиков, признание успехов и лидерства в различных видах человеческой деятельности – науке, бизнесе, спорте и т.д. Учитывая сильное влияние рейтингов на различные аспекты жизни, актуальным является понимание того, как работают соответствующие методы ранжирования, каковы их основные допущения и ограничения. Недостатки рейтингов, такие как предвзятость, которую они в себе могут заключать, могут перевесить преимущества, которые они приносят.

Важный класс алгоритмов ранжирования основан на сетевом представлении входных данных [1]. Идеи использования структуры сети для определения статуса узла уходят корнями в социальные науки, где для этого были предложены различные метрики *центральности*. Реальные приложения сетевых алгоритмов ранжирования охватывают широкий спектр задач, включая разработку поисковых систем и систем рекомендаций, оценку научных исследований и идентификацию влиятельных распространителей (информации, фейков, инфекций, инноваций), многие другие.

Хотя методы ранжирования в сетях широко обсуждаются последние 30-40 лет (особенно в связи с ранжированием в поисковых системах WWW), только в последнее время начало обсуждаться влияние на свойства и эффективность алгоритмов ранжирования эволюции сети. Сети описывают системы, которые, как правило, растут добавлением новых узлов и связей. В WWW, например, постоянно добавляются новые сайты с гиперссылками на страницы уже существующие, а еще происходит обновление. К каким искажениям рейтингов узлов по полезности, значимости, пригодности и т.п. может приводить представление структуры сети как статической? Насколько значимым при ранжировании является понимание времени-зависимости сети? Каким образом можно учесть при ранжировании время-зависимость и насколько это изменяет рейтинги? В данной работе проведено исследование эффектов учета зависимости от времени в системах ранжирования, основанных на центральностях узлов в сложных сетях, которое в некоторой степени может найти ответы на эти вопросы.

1. Подходы к ранжированию в сложных сетях

Задача *ранжирования* предполагает присвоение рангов неким объектам исследуемой системы, что позволяет отсортировать, упорядочить эти объекты по важности, полезности или релевантности. В зависимости от вида и характера задачи для ранжирования применяются разные методы. Самый простой случай, когда ранжирование происходит на основании прямо приписанных элементам характеристик – измеренных или вычисленных. Например, упорядочить людей по росту или возрасту, или слова некоторого текста по частоте встречаемости. Все усложняется, когда значимость элементов системы определяется множеством взаимодействий и взаимозависимостей, как например, авторитетность в коллективе, влияние научной публикации или порядковый номер сайта в отчете поисковой системы. Для такого рода систем удобным оказывается описание их как *сетей*, в качестве *узлов* которых выступают элементы системы, а взаимодействиям соответствуют *связи* между узлами (*ребра*) [1]. Сетевые алгоритмы ранжирования используют сетевые представления данных для вычисления внутренней ценности, релевантности или важности отдельных узлов в системе. Идея использования структуры сети для определения статуса узла давно развивается в социальных науках, где показатели ранжирования на уровне узла обычно называются метриками *центральности*. Применимость того или иного алгоритма ранжирования, как правило связана с типом сети.

Базовые типы и классы сетей

При выборе и вычислении показателей центральности существенно, является ли сеть простой неориентированной или ориентированной. Пример, *простой* неориентированная сеть социальная сеть, в которой узлы – люди, а ребра – родственные отношения между ними. В *ориентированной* сети учитывается направленность связей между узлами. Например, WWW, где узлы – это сайты, а связи – гиперссылки с одного сайта на другой.

Полную информацию для всех этих типов сетей, как правило, кодируют матрицей смежности A размером $N \times N$, где N – количество узлов: $A_{ij} = 1$, если ребро связывает узел i с узлом j , и $A_{ij} = 0$ в противном случае. Для простой сети матрица смежности симметрична, а для ориентированной ребро идет от узла i к узлу j .

При моделировании реальных систем выделяют два класса сетей, существенно различающихся своими свойствами и характеристиками. Это однородные сети и неоднородные. Определяются эти классы формой распределения степеней.

Степень узла – это количество ребер, связанных с узлом. Степень k_i узла i в сети, описываемой матрицей смежности A , может быть вычислена как $k_i = \sum_{j=1}^N A_{ij}$. Распределение степеней $P(k)$ описывает, с какой вероятностью случайно выбранная вершина имеет степень k .

Для *однородных* сетей степени вершин незначительно отличаются от средней степени. Примером такой сети является случайный граф Эрдеша-Реньи [1], где два узла связаны с некоторой вероятностью p , и распределение степеней – распределение Пуассона $P(k) \cong e^{-pN} \frac{(pN)^k}{k!}$. Для *неоднородных* сетей характерен большой разброс степеней вершин (несколько порядков по величине). Подавляющее большинство реальных сетей характеризуются степенным распределением степеней $P(k) \sim k^{-\gamma}$ со значением показателя γ между 2 и 3, такие сети называют *масштабно-инвариантными* или *безмасштабными*. Свойства таких сетей во многом определяются наличием небольшого числа вершин очень большой (по сравнению со средней) степени, хабов. Именно такие сети, проявляющие нетривиальные топологические свойства, называют сложными [1].

Статические показатели центральности

Ранжирование в сетях направлено на выявление наиболее значимых и влиятельных узлов. Первыми были предложены метрики центральности, учитывающие только топологию сети, закодированную в ее матрице смежности A (называемые статическими сетевыми *метриками центральности*), такие как степень, близость, посредничество, *PageRank*, центральность собственного вектора, авторитетность, центральность хаба, центральность k -оболочки [2].

Степень узла в качестве показателя центральности предполагает, что узел важен, если у него много связей. В ориентированных сетях степень узла делится на степень захода (*in-degree*) и степень исхода (*out-degree*), $k_i^{in} = \sum_{j=1}^N A_{ij}$ и $k_i^{out} = \sum_{j=1}^N A_{ji}$, то есть количество входящих и исходящих ребер соответственно.

Центральность близости (Closeness centrality) определяется как $c_i = (N - 1) / \sum_{j \neq i} d_{ij}$, где d_{ij} – это расстояние между узлами i и j . Узел в этом контексте является центральным, если он «близок» (в сетевом смысле) ко многим другим узлам. Поскольку для узлов, недостижимых из i , d_{ij} полагают равным бесконечности, c_i в таком случае оказывается равным 0. Для несвязных сетей поэтому переходят к определению центральности узла i как гармонического среднего его расстояния до других узлов [2]: $c_i = \left(\sum_{j \neq i} \frac{1}{d_{ij}} \right) / (N - 1)$.

Центральность посредничества (Betweenness centrality) – это показатель, измеряющий сколько раз через узел проходят все возможные кратчайшие пути. Может быть вычислен как $bc_i = \frac{1}{(N-1)(N-2)} \sum_{s \neq i \neq t} \frac{\sigma_{st}^{(i)}}{\sigma_{st}}$, где σ_{st} – общее количество кратчайших путей из s в t , а $\sigma_{st}^{(i)}$ – это число кратчайших путей, которые содержат i [2].

Центральность собственного вектора (Eigenvector centrality) – это метод определения центральности, предполагающий, что узел важен, если он соединен с другими важными узлами и поэтому получает более высокое значение центральности. Для его вычисления изначально принимают

одинаковую оценку $s_i^{(0)} = 1$ для всех узлов i . На каждом последующем шаге оценки итеративно обновляют в соответствии с уравнением $s_i^{(n+1)} = \sum_j A_{ij} s_i^{(n)}$.

Центральность Каца (Katz centrality) – показатель, как и центральность собственного вектора, основанный на том, что узел важен, если он связан с другими важными узлами, – но, в отличие от центральности собственного вектора, он присваивает каждому узлу определенное минимальное значение, тем самым применим к направленным ациклическим графам. Центральность по Кацу может давать неудовлетворительные результаты в сетях с неоднородным распределением степеней исхода. В [3] описано применение этого показателя в системе подсчета очков для ранжирования в видах спорта на основе матчей один на один. В терминах сложных сетей каждый игрок i – это узел, а вес направленного ребра $j \rightarrow i$ – это количество побед игрока j над игроком i . Основная идея сетевой системы подсчета выигрышей и поражений заключается в том, что игрок силен, если он способен победить других сильных игроков, то есть игроков, которые победили многих соперников.

Метрика центральности *PageRank* была введена для ранжирования веб-страниц в WWW. Для направленной сети вектор s оценок *PageRank* определяется как $s = \alpha P s + (1 - \alpha) v$, где $P_{ij} = A_{ij} / k_j^{out}$ – матрица перехода сети, v – вектор телепортации, а α – параметром телепортации. Как и в случае центральности Каца, оценка *PageRank* узла определяется всеми путями, проходящими через этот узел. Параметр телепортации α управляет экспоненциальным затуханием вкладов более длинных путей. При малых α (но > 0) длинные пути дают незначительный вклад, и, если флуктуации исходящей степени достаточно малы, ранжирование по *PageRank* приблизительно сводится к ранжированию по входящей степени. Когда α близко к единице (но < 1), длинные пути дают существенный вклад в оценку узла. Считается, что алгоритм *PageRank* хорошо описывает реальное поведение в сети пользователя, просматривающего интернет как переходя по гиперссылкам, размещенным на веб-страницах, так и – возможно, когда становится скучно или он застревает – повторными запусками поиска (совершая «телепортацию»). *PageRank* зависит от того, насколько узел доступен для других узлов.

Алгоритм *HITS (Hyperlink-Induced Topic Search)* – популярный метод ранжирования на основе собственных векторов, изначально предназначенный для ранжирования веб-страниц в WWW. В однодольной ориентированной сети алгоритм *HITS* присваивает две оценки каждому узлу самосогласованным образом. Показатель h , называемый оценкой *хаб-центральности* узла, является большим для узлов, которые указывают на множество авторитетных узлов. Другой показатель a , называемый показателем *авторитетности*, является высоким для узлов, на которые указывает множество хабов [2]. Соответствующие уравнения для оценок узлов a и h имеют вид: $a = \alpha A h$, $h = \beta A^T a$, где α и β – параметры метода. В контексте сетей цитирования, например, тематические обзоры идентифицируются как хабы, поскольку они содержат много ссылок на влиятельные статьи.

Темпоральные (время-зависимые) сети и алгоритмы ранжирования, зависящие от времени

Многие системы изменяются со временем, и, как правило, эволюция системы отражается на показателях центральности, которые вычисляются на основе статической сети. Ранжирование на основе статических показателей центральности может приводить к существенным искажениям. Узлы могут получить низкие или высокие оценки в основном из-за времени, когда они вошли в систему, что очень нежелательно, когда мы заинтересованы в выявлении внутренней пригодности или качества узлов. Например, статья, опубликованная несколько месяцев назад, обречена на плохой рейтинг по количеству цитирований, потому что у нее еще не было времени показать свой потенциал, привлекая соответствующее количество цитирований.

Темпоральная или время-зависимая сеть – это сеть, в которой тем или иным образом учитывается время. Сделать это можно по-разному. Информация о времени может выступать в качестве атрибута узлов или ребер. Например, в сети цитирования научных статей это может быть дата публикации. Альтернативно, можно картировать сеть через определенные интервалы времени, тем самым получая информацию о изменениях в ее топологии.

Предполагая, что сеть невзвешенная, алгоритмы ранжирования, зависящие от времени, принимают в качестве входных данных матрицу смежности $A(t)$ сети в данный момент времени t и информацию о возрасте узла и/или отметке времени на ребре. Их можно разделить на три категории:

(I) перемасштабирование (рескейлинг) оценок на основе оценок (центральностей) узлов; (II) метрики, которые явно «штрафуют» старые узлы или ребра; (III) метрики, основанные на моделях роста сети, которые предполагают существование скрытого параметра пригодности, представляющего собой показатель успешности узла в системе.

Перемасштабируемые по времени метрики центральностей узлов

Если интерес представляет внутренняя пригодность узлов, а для эволюционирующей сети показатели центральности могут быть сильно искажены возрастом узла, то можно просто взять исходные статические оценки центральности в качестве входных данных и перемасштабировать их, сравнивая оценку каждого узла только с оценками узлов близкого возраста.

Пример применения *перемасштабирования центральности степени* захода – сети цитирования. Смещение числа цитирований статьи устраняется путем оценки среднего значения $\mu_i(c)$ и стандартного отклонения $\sigma_i(c)$ количества цитирований для статей, опубликованных примерно в то же время, что и статья i . Затем систематическая зависимость c_i от времени публикации устраняется вычислением перемасштабированного количества цитирований

$$R_i(c) = \frac{c_i - \mu_i(c)}{\sigma_i(c)}. \quad (1)$$

$R_i(c)$ – оценка статьи i в пределах ее *окна оценивания*. Значения $R(c)$ больше или меньше нуля указывают, превосходит ли статья статьи аналогичного возраста или отстает от них. Не существует принципиального критерия для выбора статей, «опубликованных в аналогичное время», по которым вычисляются $\mu_i(c)$ и $\sigma_i(c)$. В случае данных цитирования это вычисление ограничивают фиксированным числом Δ статей, опубликованных непосредственно до и непосредственно после статьи i , или взвешивают оценку статей с помощью гауссовой функции, центрированной на статье i . Для сетей цитирования вычисляют и *перемасштабированный PageRank*. Временное смещение оценки подавляют с помощью уравнения

$$R_i(p) = \frac{p_i - \mu_i(p)}{\sigma_i(p)}, \quad (2)$$

аналогично перемасштабированию степени. Предполагая, что узлы помечены в порядке убывания возраста, μ_i и σ_i вычисляются по «движущемуся» временному окну $[i - \Delta/2, i + \Delta/2]$, центрированному на статье i .

В принципе, любую статическую метрику центральности можно перемасштабировать подобным образом. Единственный параметр перемасштабированной оценки – размер Δ временного окна, используемого для расчета μ_i и σ_i . Слишком большие значения Δ приводят к ранжированию, смещенному в сторону старых узлов, а слишком маленькие – к ранжированию, сильно подверженному статистическим колебаниям: узел, случайно имеющий только статьи с низкой степенью в своем небольшом временном окне, будет иметь высокую оценку.

Метрики, явно штрафующие старые узлы или ребра

Многие метрики, которые явно штрафуют старые узлы и ребра, основаны на алгоритме *PageRank*. В отличие от перемасштабируемых показателей, которые апостериорно нормализуют оценки с помощью статических метрик центральности, эти показатели напрямую включают информацию о времени в расчет. Функции старения вершин и ребер могут быть очень разными для разных реальных систем, поэтому метрики, явно штрафующие узлы или ребра, в данной работе не рассматривались.

Ранжирование, основанное на моделях роста сети

Если есть математическая модель, хорошо описывающая развитие сети, то оценить важность узла можно, подгоняя модель под имеющиеся данные.

Примером такого подхода является ранжирование на основе модели предпочтительного присоединения, где каждый узел наделен параметром пригодности, представляющим важность узла, воспринимаемую другими узлами сети, – модели Бианкони-Барабаси. Пригодность узла представляет собой присущую узлам способность приобретать новые входящие связи. В версии модели *пригодности* (или фитнес-модели) с направленной сетью вероятность $\Pi_i(t)$ того, что новая входящая связь будет получена узлом i в момент времени t , определяется как $\Pi_i(t) \sim (k_i^{in}(t) + C)\eta_i$, где η_i – пригодность узла i , а C – аддитивная константа. Обозначая как t_i время появления узла i , это правило присоединения

приводит к следующему уравнению для зависимости ожидаемой степени узла от времени $\langle k_i^{in}(t) \rangle = m \left(\frac{t}{t_i} \right)^{\beta(\eta)}$, где, если распределение пригодности ограничено, $\beta(\eta)$ является решением самосогласованного уравнения вида $\beta(\eta) = \eta / C[\beta(\eta)]$, а $C[\beta]$ зависит только от выбранного распределения пригодности. Используя это уравнение, можно оценить пригодность узла в реальных данных. Для этого достаточно проследить эволюцию степени $k^{in}(t)$ узлов и правильно подогнать $k^{in}(t)$ под степенной закон, чтобы получить показатель β , который пропорционален пригодности узла.

На основании сделанного обзора показателей и алгоритмов ранжирования в сетях было принято решения для дальнейшего исследования в первую очередь рассмотреть следующий набор показателей:

- Степень, чтобы выделить узлы как важные узлы с наибольшим количеством связей.
- Близость, чтобы выделить узлы, которые способны достигнуть остальные узлы преодолев минимальное расстояние.
- Посредничество, чтобы определить через какие узлы проходит наибольшее количество кратчайших путей.
- Собственный вектор, чтобы учесть важность соседних узлов.
- *PageRank*, так как в ориентированных сетях использование этой центральности позволяет учесть качество и количество входящих ссылок.
- Перемасштабированные степень и *PageRank*.

Выбор обусловлен во многом практически универсальным характером этих показателей. Они в меньшей степени специализированы под конкретные задачи ранжирования в конкретных сетях.

2. План исследований

Основная рассматриваемая гипотеза. Ранжирование должно выявить наиболее значимые (и упорядочить по значимости) узлы сети. Реальные сети, как правило, эволюционируют. Внутренние свойства узлов (их значимость или пригодность) влияют на эволюцию сети (например, более пригодные узлы собирают больше связей, ссылок в процессе эволюции). С другой стороны, время поступления в сеть узла или связи влияет на топологические характеристики (очень ценный узел может *не успеть* проявиться в сети как значимый, например, гениальное изобретение, но только что сделанное). Игнорирование зависимости от времени приводит к существенному искажению в системах ранжирования, построенных на статических алгоритмах ранжирования (сопоставлять узлы разного возраста, как правило, при ранжировании не корректно). Насколько сильно это искажение? Насколько модификации показателей центральности позволяют выявлять наиболее значимые в том или ином контексте узлы?

Для оценки корреляций центральностей и других показателей использовался коэффициент ранговой корреляции Кендалла [5], вычисляемые как $r = 2(C - D)/(n(n - 1))$, где n – объем выборки, C – количество совпадений, а D – количество инверсий.

В данной работе были проведены исследования над синтетическими (сгенерированными) и реальными сетями.

Используемые модели синтетических сетей

Традиционно эксперименты проводятся, как минимум, над двумя классами синтетических сетей: однородными (Эрдеша-Реньи) и масштабно-инвариантными (Барабаси-Альберт). Первый класс сетей представляет из себя случайную сеть, где связь между узлами устанавливается с заданной вероятностью $p \in [0..1]$. Второй – это растущая сеть с *предпочтительным присоединением*, то есть в сеть в каждый момент времени поступает новый узел, и новые узлы с большей вероятностью присоединяются к узлам с большей степенью. Количество ребер m , с которыми новый узел поступает в сеть, является параметром модели. Вероятность присоединения к узлу i определяется как $P_i = k_i / \sum_j k_j$, где k_i означает степень узла i .

Стоит отметить, что модель Эрдеша-Реньи является статичной, в ней нет зависимости от времени, а сеть Барабаси-Альберт представляет эволюционирующую сеть, в ней время присутствует через порядок добавления вершин. Ни одна из этих моделей не учитывает *внутреннюю пригодность* или

значимость узлов, которую должно выявлять ранжирование. Внутренняя пригодность присутствует в расширении модели Барабаси-Альберт, так называемой фитнес-модели сети Бьянкони-Барабаси. В этой модели правило предпочтительного присоединения кроме степени узлов учитывает приписанные узлам при «рождении» значения фитнес-параметров η (пригодностей), распределенные в соответствии с некоторым законом: $\Pi_i = \eta_i k_i / \sum_j \eta_j k_j$. Закон распределения значений фитнес-параметра может быть разным, дальше рассматривается случай равномерного распределения значений фитнес-параметра в диапазоне от 0 до 1.

Используемые реальные сетевые данные

Для эксперимента с реальной сетью использовался набор данных *email-Eu-core-temporal-Dept1* [6]. Он представляет переписку между членами отдела исследовательского института. Набор данных не содержит переписку с внешним миром. В этих данные содержатся ID получателя и отправителя, а еще время, когда было отправлено сообщение. Время измеряется в секундах от начала исследования. Данные собирались в течении 803 дней. Сеть, построенная с помощью этих данных, является взвешенной и ориентированной.

Этапы исследования

Этап 1. Для классических неориентированных и ориентированных сетей Эрдеша-Реньи и Барабаси-Альберт анализировались формы распределений базового набора центральностей, а также ранговые корреляции между ними.

Этап 2. Используя модель эволюционирующей сети Бианкони-Барабаси, в которой на топологию сети существенно влияют ассоциированные с вершинами значения пригодности, изучалось, как эти пригодности коррелируют с основными показателями центральности.

Этап 3. На этом этапе исследовалось, улучшают ли алгоритмы перемасштабирования статических центральностей качество ранжирования. Для этого для сетей, сгенерированных на этапе 2, была реализована процедура перемасштабирования основных центральностей, после чего определены их корреляции со значениями фитнес-параметра.

Ранжирование выполнялось на основе следующих показателей: степень, центральность близости, центральность посредничества, *PageRank*; центральность собственного вектора; центральность по авторитетности и хаб-центральность; для учета зависимости от времени были вычислены перемасштабированные степень и *PageRank*.

В табл. 1 приведена сложность алгоритмов для вычисления основных выбранных метрик.

Табл. 1. Вычислительная сложность алгоритмов расчета центральностей (V – количество вершин, а E – количество ребер)

	Сложность для неориентированных графов
Степень	$O(E)$
Близость	$O(VE)$
Посредничество	$O(VE)$
<i>PageRank</i>	$O(E + V \log V)$
Центральность собственного вектора	$O(V^3)$
Хаб-центральность	$O(V^3)$
Авторитетность	$O(V^3)$

Для оценок корреляции центральностей вычислялась ранговая корреляция Кендалла. В данной работе корреляция характеризовалась как *сильная*, если значение коэффициента корреляции по модулю лежит в диапазоне от 0.66 до 1, *средняя* – от 0.35 до 0.65 и *слабая* – при значениях от 0.10 до 0.34. При значениях ниже 0.10 считалось, что корреляция отсутствует.

3. Практическая часть

Инструментарий

Исследование проводилось в программной среде *RStudio* с использованием библиотеки *igraph*, предназначенной для работы с сетями. При выполнении работы были написаны независимые скрипты соответственно этапам исследования. Скрипты загружены на *GitHub* [7].

Этап 1. Исследование формы распределения и корреляций центральностей сетей Эрдеша-Реньи и Барабаси-Альберт

Случай неориентированных сетей

На первом этапе для изучения формы распределения центральностей и корреляций между ними были сгенерированы классические (простые и неориентированные) сети Эрдеша-Реньи (ЭР) и Барабаси-Альберт (БА). С целью изучения зависимостей от размера и плотности при генерации количество вершин N изменялось от 1000 до 10 000, а средняя степень $\langle k \rangle$ от 4 до 20. Ниже представлены характерные графики распределений для сетей, состоящих из 10 000 узлов с $\langle k \rangle = 8$. Для случайных сетей ЭР это соответствует вероятности связи между вершинами $p=0.0008$, а для сети БА параметру, определяющему количество узлов, к которым присоединяется новый узел, $m = 4$.

Для сетей этих двух классов были вычислены: степень k , центральность близости s , центральность по посредничеству b , *PageRank*, центральность собственного вектора, центральности авторитетности и хаба. Для этих центральностей были построены распределения (см. рис. 1), а также вычислены корреляции.

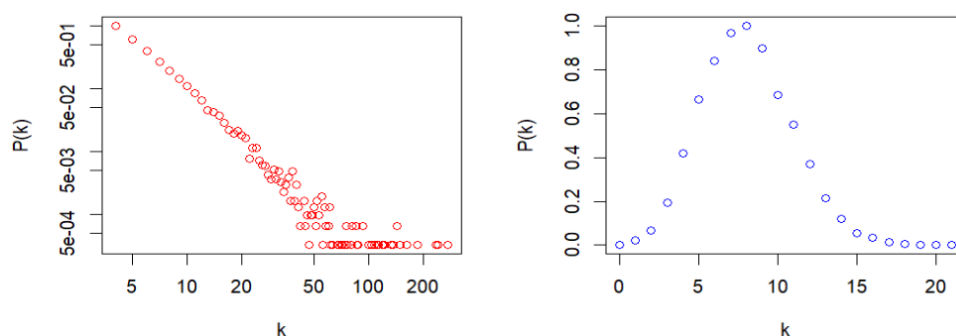


Рис. 1. Распределения степеней $P(k)$, где k – степень узла. Слева для сети Барабаси-Альберт в $\log\text{-}\log$ масштабе, справа для сети Эрдеша-Реньи. Для обеих сетей $N = 10000$, $\langle k \rangle = 8$

На рис. 1 видно, что распределение степеней сети ЭР имеет форму распределения Пуассона, эта сеть однородна в том смысле, что степени всех вершин близки к $\langle k \rangle$. В то же время $P(k)$ для сети БА описывается степенным законом $P(k) \sim k^{-\gamma}$ с $\gamma \approx 3$. Здесь максимальная степень существенно превышает $\langle k \rangle$. В таких неоднородных сетях хабы играют выделенную роль, что, естественно, влияет на их ранг.

На рис. 2 показаны распределения центральности близости. Наибольшие значения s соответствуют узлам с наименьшим суммарным расстоянием до остальных узлов. В сети Эрдеша-Реньи максимальное значение $s_{\max} \sim 0.24$ при среднем $\langle s \rangle \sim 0.21$, а для сети БА – $s_{\max} \sim 0.37$ при среднем $\langle s \rangle \sim 0.25$. Таким образом, значения центральности близости также свидетельствуют о большей неоднородности сети БА. Примечательно, что в сети БА наибольшие значения центральности близости имеют хабы, именно от них можно достичь другие узлы быстрее.

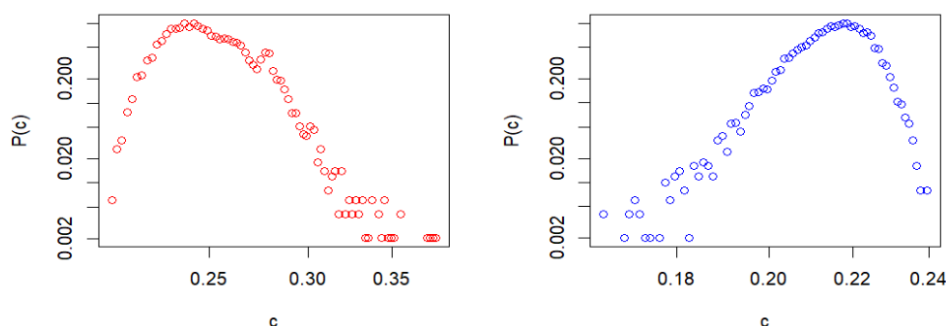


Рис. 2. Распределения центральности близости $P(c)$. Слева для сети Барабаси-Альберт, справа для сети Эрдеша-Реньи. Для обеих сетей $N = 10000$, $\langle k \rangle = 8$

На рис. 3 показаны распределения $P(bc)$ центральности по посредничеству. Распределение $P(bc)$ для сети БА имеет явно выраженную форму степенного распределения $P(bc) \sim bc^{-\alpha}$, что говорит о наличии малого числа узлов, которые встречаются на подавляющем большинстве кратчайших путей. Для сети ЭР максимальное значение посредничества более чем в 20 раз меньше, форма распределения $P(bc)$ говорит об однородности сети.

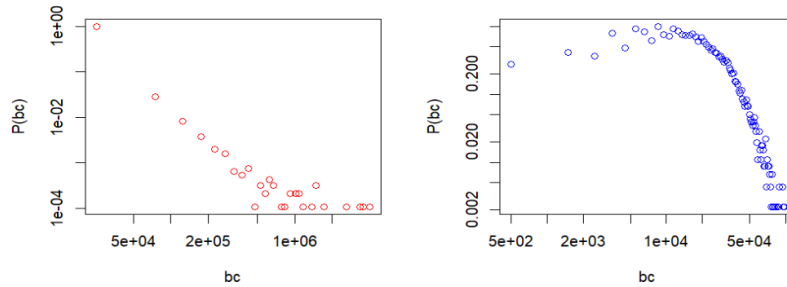


Рис. 3. Распределения центральности по посредничеству $P(bc)$ в $\log\text{-}\log$ масштабе. Слева для сети Барабаси-Альберт, справа для сети Эрдеша-Реньи. Для обеих сетей $N = 10000$, $\langle k \rangle = 8$

На рис. 4 показаны распределения $P(pr)$ центральности PageRank. Для сети ЭР форма распределения близка к пуассоновскому распределению, практически отсутствуют узлы с высокими показателями PageRank (по сравнению с сетью БА). Для сети БА распределение описывается степенным законом, т.е. присутствует небольшое число узлов с высокими значениями PageRank, «высокоавторитетных» хабов, доминирующих во всей сети.

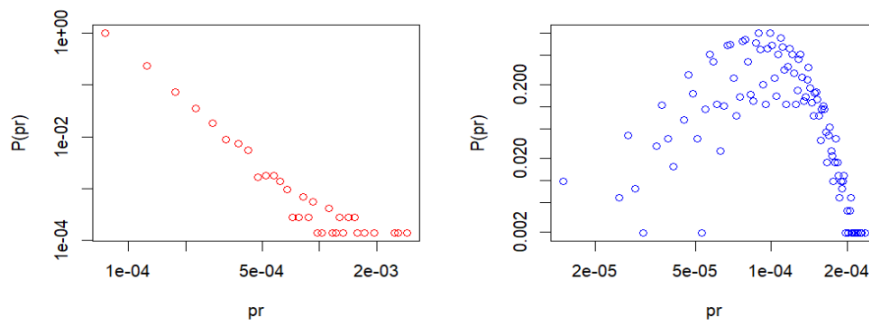


Рис. 4. Распределения $P(pr)$ центральности PageRank в $\log\text{-}\log$ масштабе. Слева для сети Барабаси-Альберт, справа для сети Эрдеша-Реньи. Для обеих сетей $N = 10000$, $\langle k \rangle = 8$

На рис. 5 показаны распределения по центральности собственного вектора, авторитетности и хаб-центральности.

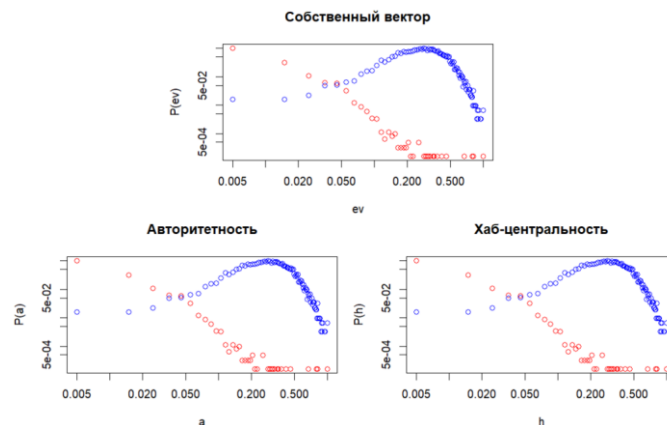


Рис. 5. Распределения центральностей собственного вектора $P(ev)$, авторитетности $P(a)$ и хаб-центральности $P(h)$ для сетей Барабаси-Альберт (красные круги) и сетей Эрдеша-Реньи (синие).
Для обеих сетей $N = 10000$, $\langle k \rangle = 8$

Как и ожидалось, для неориентированных и связных сетей они полностью идентичны. Для хаб-центральности и авторитетности это объясняется тем, что авторитетность учитывает степени захода, а хаб-центральность – степени исхода, а так как здесь рассматриваются неориентированные сети, эти центральности получают одинаковые значения для каждого узла.

Таким образом, для однородных сетей распределения практически всех центральностей отражают эту однородность. Все узлы «равноправны» и равно значимы, любое ранжирование будет фактически отражать статистические флуктуации. Для масштабно-инвариантных сетей, таких как сети БА, широкомасштабные распределения для степени, центральности посредничества, *PageRank*, центральности собственного вектора, указывают на выделенность небольшого набора узлов (хабов).

Для оценки отношений между ранжированиями по этим центральностям, для понимания того, насколько они «работают» независимо, были вычислены коэффициенты ранговой корреляции Кендалла. Они представлены в табл. 2 и 3 для сетей ЭР и БА, соответственно.

Табл. 2. Корреляции Кендалла центральностей для сети Эрдеша-Реньи

	<i>Degree</i>	<i>Closeness</i>	<i>Betweenness</i>	<i>PageRank</i>	<i>Eigenvector</i>	<i>Authority</i>	<i>Hub</i>
<i>Degree</i>	1.00	0.81	0.92	0.95	0.79	0.79	0.79
<i>Closeness</i>	0.81	1.00	0.78	0.70	0.91	0.91	0.91
<i>Betweenness</i>	0.92	0.78	1.00	0.89	0.74	0.74	0.74
<i>PageRank</i>	0.95	0.70	0.89	1.00	0.68	0.68	0.68
<i>Eigenvector</i>	0.79	0.91	0.74	0.68	1.00	1.00	1.00
<i>Authority</i>	0.79	0.91	0.74	0.68	1.00	1.00	1.00
<i>Hub</i>	0.79	0.91	0.74	0.68	1.00	1.00	1.00

Табл. 3. Корреляции Кендалла центральностей для сети Барабаси-Альберт

	<i>Degree</i>	<i>Closeness</i>	<i>Betweenness</i>	<i>PageRank</i>	<i>Eigenvector</i>	<i>Authority</i>	<i>Hub</i>
<i>Degree</i>	1.00	0.42	0.73	0.92	0.39	0.39	0.39
<i>Closeness</i>	0.42	1.00	0.62	0.30	0.91	0.91	0.91
<i>Betweenness</i>	0.73	0.62	1.00	0.61	0.58	0.58	0.58
<i>PageRank</i>	0.92	0.30	0.61	1.00	0.27	0.27	0.27
<i>Eigenvector</i>	0.39	0.91	0.58	0.27	1.00	1.00	1.00
<i>Authority</i>	0.39	0.91	0.58	0.27	1.00	1.00	1.00
<i>Hub</i>	0.39	0.91	0.58	0.27	1.00	1.00	1.00

По данным из табл. 2 можно заметить, что значения ранговой корреляции Кендалла для сети Эрдеша-Реньи могут быть названы сильными. Для центральности собственного вектора, хаб-центральности и авторитетности корреляция равна 1, их значения полностью совпадают, поэтому в случае неориентированных сетей достаточно вычислять и анализировать только одну из этих трех центральностей. Таким образом, что в случае ранжирования узлов однородных сетей, похожих на сети Эрдеша-Реньи, результаты слабо зависят от выбора показателя центральности для ранжирования (определения рейтинга). Можно выбрать самый простой с точки зрения вычислительной сложности – центральность степени.

Иная картина – для неоднородных сетей. Табл. 3 показывает, что сильную ранговую корреляцию можно наблюдать между степенью и посредничеством, степенью и *PageRank*, близостью и собственным вектором. При этом нельзя ничего сказать о том, какая же из центральностей в случае неоднородных сетей укажет на наиболее пригодные, авторитетные и влиятельные узлы. Во многом это определяется задачами ранжирования.

Случай ориентированных сетей

Были сгенерированы ориентированные сети Эрдеша-Реньи и Барабаси-Альберт, поэтому рассматривались отдельно степени исхода и захода. Ниже в таблицах использованы следующие обозначения: *In_d* – центральность по степени захода; *Out_d* – центральность по степени исхода;

C – центральность близости; Bc – центральность по посредничеству; Pr – *PagRank*; Ev – центральность собственного вектора; A – центральность по авторитетности.

Характер распределений центральностей для ориентированных сетей БА и ЭР остался таким же, как и для неориентированных, поэтому они здесь не приведены. Изменились только средние значения некоторых центральностей. Отметим, что распределение центральности по степени исхода и хаб-центральности для сети БА не имеет смысла рассматривать, так как обе центральности зависят от степени исхода, одинаковой для практически всех узлов сети БА.

Результаты вычисления ранговой корреляции Кендалла в сети ЭР показаны в табл. 4. Для большинства центральностей корреляция осталась сильной, что обусловлено высокой степенью однородности сети.

Табл. 4. Ранговые корреляции Кендалла центральностей для ориентированной сети ЭР

	In_d	Out_d	C	Bc	Pr	Ev	A	Hub
In_d	1.00	-0.01	0.50	0.50	0.73	0.50	0.80	-0.01
Out_d	-0.01	1.00	0.50	0.49	-0.01	0.50	-0.01	0.80
C	0.50	0.50	1.00	0.76	0.40	0.97	0.48	0.47
Bc	0.50	0.49	0.76	1.00	0.47	0.76	0.42	0.41
Pr	0.73	-0.01	0.40	0.47	1.00	0.40	0.53	-0.01
Ev	0.50	0.50	0.97	0.76	0.40	1.00	0.48	0.48
A	0.80	-0.01	0.48	0.42	0.53	0.48	1.00	-0.01
Hub	-0.01	0.80	0.47	0.41	-0.01	0.48	-0.01	1.00

Для сети Барабаси-Альберт сильную ранговую корреляцию можно наблюдать между авторитетностью и степенью захода, хаб-центральностью и степенью исхода, посредничеством и близостью, а также между посредничеством и собственным вектором (см. табл. 5).

Табл. 5. Ранговые корреляции Кендалла центральностей для ориентированной сети БА

	In_d	Out_d	C	Bc	Pr	Ev	A	Hub
In_d	1.00	-0.03	0.18	0.90	0.98	0.19	0.91	0.14
Out_d	-0.03	1.00	-0.03	-0.02	-0.03	-0.03	-0.03	-0.01
C	0.18	-0.03	1.00	0.13	0.18	0.93	0.18	0.89
Bc	0.90	-0.02	0.13	1.00	0.90	0.14	0.85	0.10
Pr	0.98	-0.03	0.18	0.90	1.00	0.19	0.90	0.14
Ev	0.19	-0.03	0.93	0.14	0.19	1.00	0.19	0.92
A	0.91	-0.03	0.18	0.85	0.90	0.19	1.00	0.14
Hub	0.14	-0.01	0.89	0.10	0.14	0.92	0.14	1.00

Таким образом, на данном этапе было выявлено, что размер сети (количество узлов) и плотность сети (количество связей, определяемое вероятностью соединения в сети ЭР и параметром m в сети БА) практически не влияют на характер распределения центральностей, но изменяют величину центральностей. Как и ожидалось, авторитетность и хаб-центральности совпадают в простых неориентированных сетях, и с этими двумя центральностями полностью совпадает собственный вектор, что делает эти три центральности взаимозаменяемыми, хотя в ориентированных сетях они различимы.

Центральности позволяют выполнять ранжирование узлов, но, опираясь на классические модели сетей, нельзя понять, насколько при таком ранжировании определяются самые пригодные, полезные, авторитетные, значимые узлы? Это, естественно, зависит от рассматриваемой системы и решаемых задач. Для эволюционирующих сетей, таких, например, как сеть Барабаси-Альберт, топологические характеристики связаны с временем поступления узлов в сеть. Примером является «эффект первопроходца» [1] – степень вершины пропорциональна $\sqrt{t}$, где t – время жизни узла в сети, т.е. вне зависимости от внутренних свойств узла, его степень и коррелирующие с ней центральности, а значит и соответствующий ранг, определяется моментом его поступления в систему. Насколько сильно влияет время на ранжирование? Для изучения этого можно использовать модель Бианкони-Барабаси [1], в которой узлам эволюционирующей сети приписывается некоторое значение пригодности (фитнесс-параметр).

Этап 2. Исследование ранжирования в сетях Бианкони-Барабаси с предопределенными значениями пригодности узлов

Необходимо понять, как соотносятся основные центральности неориентированных и ориентированных сетей с пригодностями узлов, если эти пригодности влияют на эволюцию сети.

Рассматривались сети Бианкони-Барабаси (ББ) с теми же параметрами, что и для сетей БА на этапе 1. Значения фитнес-параметра в этом исследовании были сгенерированы случайным образом в диапазоне от 0 до 1 (равномерное распределение). В [1] показано, что в отличие от сети БА, где распределение степеней описывается степенным законом $P(k) \sim k^{-\gamma}$ с $\gamma \approx 3$, для сети ББ с равномерным распределением значений фитнес-параметра $P(k) \sim \frac{k^{-(1+C^*)}}{\ln k}$, где $C^*=1.255$ (приблизительно, степенной закон с $\gamma \approx 2.255$).

Отметим, что распределения центральностей сети ББ похожи на распределения для сети БА, поэтому здесь приведены. Внимание было сконцентрировано на корреляции фитнес-параметра с центральностями. В табл. 6 показаны результаты вычислений ранговых корреляций Кендалла между центральностями и значениями фитнес-параметра для неориентированной сети ББ. Как можно заметить, корреляции центральностей не сильно отличаются от значений для сети БА (см. табл. 3). Корреляции значений фитнес-параметра со степенью, посредничеством и PageRank заметно выражены, но все равно достаточно слабые.

Табл. 6. Ранговые корреляции Кендалла центральностей и фитнес-параметра для неориентированной сети Бианкони-Барабаси

	Degree	Closeness	Betweenness	PageRank	Eigenvector	Fitness
Degree	1.00	0.31	0.61	0.87	0.29	0.30
Closeness	0.31	1.00	0.59	0.16	0.90	0.03
Betweenness	0.61	0.59	1.00	0.45	0.56	0.18
PageRank	0.87	0.16	0.45	1.00	0.14	0.30
Eigenvector	0.29	0.90	0.56	0.14	1.00	0.02
Fitness	0.30	0.03	0.18	0.30	0.02	1.00

В табл. 7 показаны результаты вычислений ранговых корреляций для ориентированной сети ББ. Как и в случае неориентированной сети, имеют место слабые корреляции значений фитнес-параметра с центральностями по степени, посредничеству и PageRank. Для направленной сети в этот набор еще добавилась авторитетность. Существенно, что фитнес коррелирует со степенью только по заходам.

Табл. 7. Ранговые корреляции Кендалла центральностей и фитнес-параметра для ориентированной сети Бианкони-Барабаси

	In_d	Out_d	C	Bc	Pr	Ev	A	Hub	Fit
In_d	1.00	-0.03	0.23	0.86	0.96	0.24	0.86	0.15	0.31
Out_d	-0.03	1.00	-0.03	-0.02	-0.03	-0.03	-0.03	-0.03	-0.01
C	0.23	-0.03	1.00	0.17	0.24	0.89	0.23	0.82	0.03
Bc	0.86	-0.02	0.17	1.00	0.87	0.17	0.79	0.10	0.29
Pr	0.96	-0.03	0.24	0.87	1.00	0.24	0.84	0.16	0.29
Ev	0.24	-0.03	0.89	0.17	0.24	1.00	0.24	0.81	0.03
A	0.86	-0.03	0.23	0.79	0.84	0.24	1.00	0.15	0.28
Hub	0.15	-0.03	0.82	0.10	0.16	0.81	0.15	1.00	-0.01
Fit	0.31	-0.01	0.03	0.29	0.29	0.03	0.28	-0.01	1.00

Таким образом, корреляции значений фитнес-параметра узлов и центральностей указывают, что на то, что центральности степени исхода, близости, собственного вектора и хаба для сети ББ практически не отражают внутренней пригодности узлов. С слабой степени фитнес-параметр коррелирует с центральностями степени (захода), посредничества, авторитетности и PageRank.

Этап 3. Рескейлинг по центральностям

Как уже отмечалось, многие системы растут со временем, и естественно, этот рост оставляет отпечаток на показателях, вычисленных на статических снимках этих систем. Этот отпечаток часто принимает форму временного смещения, которое может существенно влиять на результаты,

полученные с помощью метрики: узлы могут получить низкие или высокие оценки в основном из-за времени, когда они поступили в систему. Например, недавно опубликованная статья может иметь слабый рейтинг по количеству цитирований.

Чтобы противостоять преимуществу ранних узлов в системах, где действует механизм предпочтительного присоединения, было предложено [2] количественно оценить значимость узла, вычислив число стандартных отклонений, на которое степень захода узла превышает среднюю степень захода, соответствующую времени появления узла. Эта оценка используется для выявления узлов, которые «обходят» предпочтительное присоединение и привлекают больше связей, чем можно было бы ожидать на основе их времени добавления в систему. Перемасштабированное значение центральности определяется, например, формулами (1) и (2).

Для синтетических сетей окно оценивания варьировалось. После рескейлинга значений центральностей вычислялись ранговые корреляции Кендалла. В общем случае критерия выбора оптимального значения размера окна (дельты) не существует, но для сетей Бианкони-Барабаси, когда значения фитнес-параметр закладываются и определяют динамику увеличения степеней вершин, можно определить как оптимальный такой размер окна, который максимизирует коэффициент ранговой корреляции фитнес-параметра и соответствующей перемасштабированной оценки. На рисунке 6 показана зависимость значения коэффициента корреляции Кендалла фитнес-параметра и перемасштабированных центральностей степени, посредничества и *PageRank* от размера окна. Корреляции существенно увеличились по сравнению с тем, что было до перемасштабирования (0.3 для степени и *PageRank*, 0.19 – для посредничества), но остаются далекими от сильной.

Еще один эксперимент проводился из следующих соображений. При определении рейтингов часто интерес представляют ТОП-10, или ТОП-100 или ТОП-500. В этом случае корреляции фитнес-параметра и значений центральностей становятся существенно выше, как до рескейлинга, так и после. Например, для центральности степени для всех 10000 узлов это 0.30 (при рескейлинге с окном 20 – 0.38), для ТОП-1000 – 0.62 (при рескейлинге – 0.65), для ТОП-100 – 0.73 (при рескейлинге – 0.77).

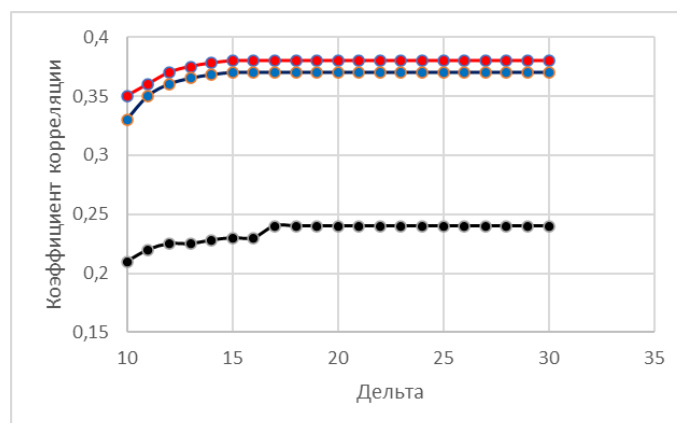


Рис. 6. Зависимости ранговой корреляции Кендалла между центральностями и фитнес-параметром для неориентированной сети Бианкони-Барабаси от размера окна (Дельта).

Корреляции показаны для степени – красным, для посредничества – черным, для *PageRank* – синим

Таким образом, эксперименты показали, что рескейлинг способен улучшить корреляцию между центральностями и фитнесом, однако вопросом остается выбор размера окна. Кроме-того, можно сказать, что и без такого учета время-зависимости узлы ТОП рейтинга будут определяться достаточно надежно.

Исследование ранжирования в реальной сети

Сеть переписок

По набору данных *email-Eu-core-temporal-Dept1* была построена сеть переписок. Сеть рассматривалась как ориентированная с временными метками у связей. Полученная сеть несвязная, и количество компонент равно 9. Это говорит о наличии групп сотрудников, ведущих переписку только внутри своей группы. В построенной сети 309 вершин, 61046 ребер, средняя степень захода (исхода) равна 197.56, диаметр – 6, среднее расстояние между узлами – 2.1.

Для этой сети были вычислены те же центральности, что и на первом этапе исследования и определены коэффициенты ранговой корреляции Кендалла (см. табл. 8). Желтым цветом обозначены корреляции, которые остались сильными для сети переписок, как и в ориентированных сетях БА и ББ (см. табл. 5 и 7). Красным цветом отмечены корреляции, которые были сильными в ориентированных сетях БА и ББ (см. табл. 5 и 7). Зеленый цвет показывает корреляции, которые стали сильными только в сети переписок.

Табл. 8. Ранговые корреляции Кендалла центральностей сети переписок

	<i>In_d</i>	<i>Out_d</i>	<i>C</i>	<i>Bc</i>	<i>Pr</i>	<i>Ev</i>	<i>A</i>	<i>Hub</i>
<i>In_d</i>	1.00	0.44	0.44	0.43	0.82	0.41	0.42	0.16
<i>Out_d</i>	0.44	1.00	0.27	0.46	0.52	0.60	0.03	0.86
<i>C</i>	0.44	0.27	1.00	0.16	0.45	-0.03	-0.08	0.03
<i>Bc</i>	0.43	0.46	0.16	1.00	0.74	0.15	-0.03	0.24
<i>Pr</i>	0.82	0.52	0.45	0.74	1.00	0.21	0.10	0.17
<i>Ev</i>	0.41	0.60	-0.03	0.15	0.21	1.00	0.72	0.74
<i>A</i>	0.42	0.03	-0.08	-0.03	0.10	0.72	1.00	0.09
<i>Hub</i>	0.16	0.86	0.03	0.24	0.17	0.74	0.09	1.00

Были также построены графики распределений центральностей (см. рис. 7). Форма распределения центральности близости похожа на Пуассоновское распределение, характерное для однородных сетей. Распределения степеней, посредничества, *PageRank*, хаб-центральностей имеют форму степенного распределения. Таким образом, рассматриваемая сеть переписок относится к неоднородным.

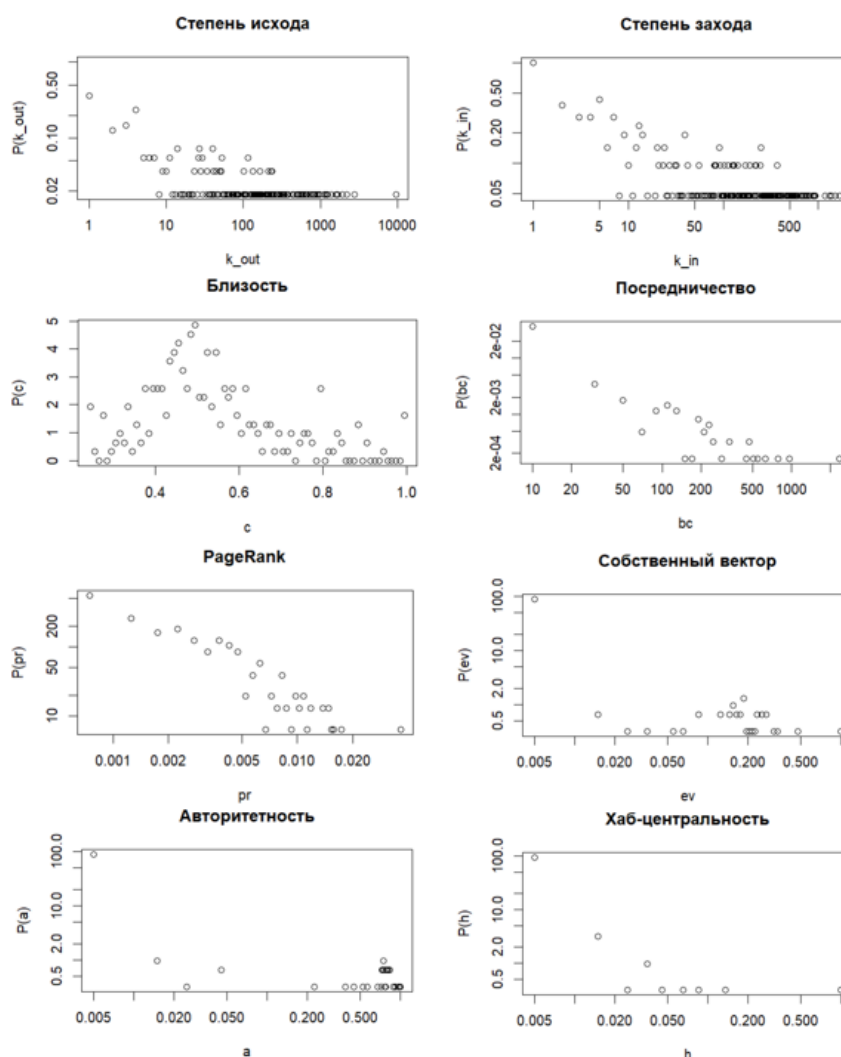


Рис. 7. Распределения центральностей, k_{out} — степень исхода, k_{in} — степень захода, c — центральность близости, bc — центральность посредничества, pr — *PageRank*, ev — центральность собственного вектора

Был проведен рескейлинг для степеней исхода и захода, а также *PageRank* при размере окна, равном 30. В табл. 9 показан ТОП-10 узлов в случае ранжирования по этим центральностям. В этой таблице столбец *Rank* – это ранг, который получил узел в ходе ранжирования. Было отобрано 10 лучших вершин по этим центральностям до и после рескейлинга, и их индекс указывался в соответствующей ячейке.

Табл. 9. Топ 10 узлов по центральностям

Rank	До перемасштабирования			После перемасштабирования		
	<i>In_Degree</i>	<i>Out_Degree</i>	<i>PageRank</i>	<i>In_Degree</i>	<i>Out_Degree</i>	<i>PageRank</i>
1	239	39	101	101	39	101
2	101	200	41	239	101	41
3	41	220	4	41	200	284
4	226	101	200	4	4	200
5	56	4	117	192	220	4
6	233	226	226	226	5	226
7	219	41	56	56	161	56
8	43	223	102	282	226	282
9	232	5	122	191	102	192
10	192	219	39	102	111	239

Примечательно, что меньше всего изменились ранги по *PageRank*, это может говорить о более высокой устойчивости алгоритма *PageRank* к временному смещению для конкретно этой сети, и что некоторые вершины имеют одинаковый ранг для центральностей по степени захода и по *PageRank*, а именно узлы под номерами: 101, 226, 56 и 282. Это означает, что эти сотрудники получают сообщения преимущественно от сотрудников, которые с точки зрения сети являются важными, и это может проявляться в сильной ранговой корреляцией для этих центральностей.

Заключение

В ходе работы исследовались эффекты учета зависимости от времени в системах ранжирования, основанных на центральностях узлов в сложных сетях. Проведенные исследования позволяют сделать следующие выводы:

- Размер и плотность сети не влияет на характер распределений центральностей.
- Для неориентированных сетей нет смысла рассматривать одновременно центральности по собственному вектору, авторитетности и хаб-центральности, так как они несут одинаковую информацию.
- Для однородных сетей не требуется рассматривать все возможные центральности, можно ограничить рассмотрение самыми легко-вычисляемыми (например, степенью).
- Характер сети (масштабно-инвариантная, пуассоновская) сильно влияет на распределения и корреляции центральностей.
- Для ранжирования с акцентом на внутренние свойства узлов (пригодность, полезность) в растущих сетях целесообразнее использовать центральности по степени и *PageRank*.
- Для уточнения ранжирования в растущих время-зависимых сетях эффект (хотя и не очень сильный) дает процедура рескейлинга показателей центральности.

В ходе выполнения работы были написаны скрипты, позволяющие воспроизвести, расширить и углубить результаты исследования. Они размещены на сайте *GitHub* по ссылке [7].

Список источников

1. Barabasi Albert-Laslo. Network Science by Albert-László Barabási. URL: <http://networksciencebook.com/> (дата обращения 7.06.2025).
2. Ranking in evolving complex networks / H. Liao, M. S. Mariani, M. Medo [et al.] // Physics Reports. 2017. Vol. 689. Pp. 1–54.

- Park J., Newman M. E. A network-based ranking system for US college football // arXive.org e-Print archive. arXiv: physics/0505169.
3. Time-aware ranking in dynamic citation networks / R. Ghosh, T.-T. Kuo, C.-N. Hsu [et al.] // 2011 IEEE 11th International Conference on Data Mining Workshops. Vancouver, 2011. DOI: 10.1109/ICDMW.2011.183.
 4. Kendall M. A New Measure of Rank Correlation // Biometrika. 1938. Vol. 30, Iss. 1-2. Pp. 81-93.
 5. Lescovec J. SNAP: Network datasets: email-Eu-core temporal network // Stanford University : [web site]. URL: <https://snap.stanford.edu/data/email-Eu-core-temporal.html> (дата обращения 04.02.2025).
 6. Ranking-in-complex-networks: [repository] // GitHub: [web platform]. – GitHub, Inc., 2025. – URL: <https://github.com/MaiqLiar/Ranking-in-temporal-networks>.