

Изучение взаимодействий между дельфинами в популяции Даутфул-Саундского пролива.

ЭАД'26

Каленбет Ульяна, Шмелева Анна

Аннотация: Данная статья посвящена изучению взаимодействий между дельфинами Даутфул-Саундского залива. В ходе работы мы найдем наиболее значимых особей в колонии, используя методы анализа сетей и графов.

Ключевые слова: сети, степенная центральность, центральность по близости, гармоническая центральность по близости, центральность по посредничеству, центральность PageRank

Описание:

Дельфины - удивительные животные, имеющие потрясающие когнитивные способности. Ученые до сих пор не изучили полностью и не расшифровали язык коммуникации дельфинов. Эти животные поражают невероятной широтой диапазона издаваемых звуков. Часть из них выполняет коммуникативную функцию, остальные используются для подводной эхолокации. Они умеют воспроизводить около шестидесяти звуковых сигналов, применяют до пяти уровней различного их комбинирования, а их так называемый словарный запас имеет объем, примерно, 14 тыс. сигналов.

Наша команда ставит перед собой задачу проанализировать сеть социальных связей дельфинов, так как изучение взаимодействий между особями несомненно поможет разобраться в том, как распространяется информация между этими животными, а такие выводы могут стать подспорьем для расшифровки языка коммуникаций дельфинов. Наша цель - построить цепочки взаимодействий между особями, а также определить, какие дельфины окажутся наиболее подходящими для проведения исследований разного типа в этой популяции.

Источники и описание данных:

Для изучения этой проблемы мы взяли данные с [сайта](#) заслуженного профессора физики Мичиганского университета Марка Ньюана, работающего в центре изучения комплексных систем и в департаменте физики. Известен своим фундаментальным вкладом в области сложных сетей и комплексных систем, за который был удостоен премии Лагранжа в 2014 году. Все данные на сайте, собранные ученым, открыты для пользователей. Выбранная нами сеть была создана группой исследователей: D. Lusseau, K. Schneider, O. J. Boisseau, P. Haase, E. Slooten и S. M. Dawson. Некоторые выводы и данные, сделанные в процессе изучения этой сети, были также опубликованы в научном журнале "Behavioral Ecology and Sociobiology" в 2003 году.

Теперь о самой сети. Она показывает частые социальные взаимодействия между дельфинами в популяции, живущей в Даутфул-Саундском проливе (Doubtful Sound) в Новой Зеландии. На момент изучения в ней состояло 62 дельфина. Вершинами графа являются конкретные дельфины, при этом ребро прокладывается между вершинами, если между

дельфинами есть частое социальное взаимодействие. Граф неориентированный, так как взаимодействие между дельфинами - это двусторонний процесс. Граф также является невзвешенным, так как ученые дополнительно не изучали интенсивность связей (для них представлял интерес лишь факт частого социального взаимодействия). Содержит всего 159 ребер, их нумерация начинается с 0. У каждой вершины, кроме id (от 0 до 61), есть уникальный Label - имя конкретного дельфина

Методология:

В рамках проведения исследования мы руководствовались материалом курса "Инструментальные методы цифровой экономики" международного центра анализа и выбора решений НИУ ВШЭ. Наибольшее внимание было уделено материалу лекции по теме "Сети".

Мы рассмотрели несколько показателей (степенная центральность, центральность по близости, гармоническая центральность по близости, центральность по посредничеству, центральность PageRank) и посчитали коэффициенты для каждого узла. Затем рассмотрели графы взаимодействий и графики распределений вершин, проанализировали полученные результаты и сформировали выводы.

Задачи:

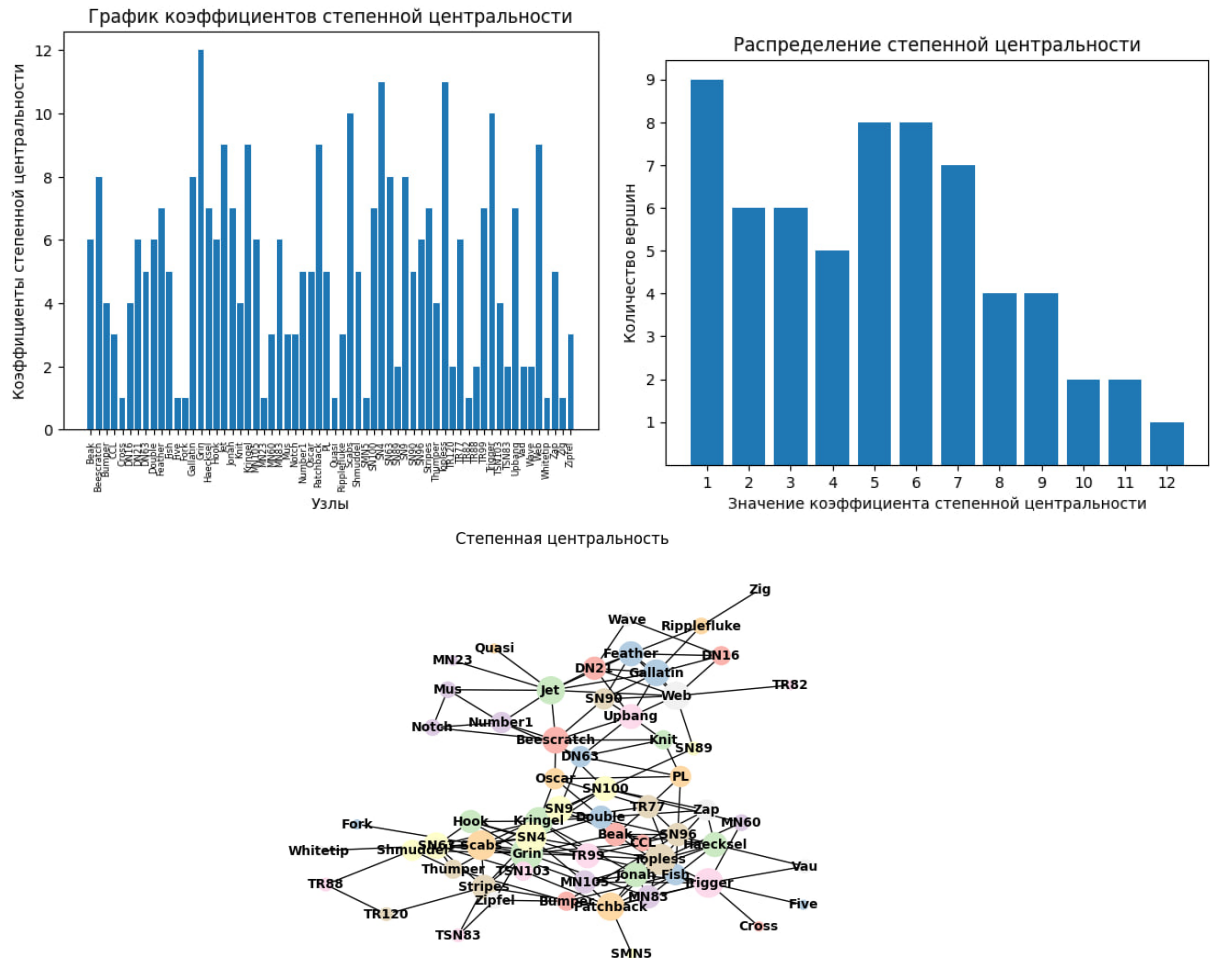
1. Выгрузка данных
2. Обработка данных с помощью библиотек Python
 - (a) pandas
 - (b) networkx
 - (c) matplotlib.pyplot
 - (d) functools
3. Подсчет и анализ ключевых показателей
 - (a) Степенная центральность
 - (b) Центральность по близости
 - (c) Гармоническая центральность по близости
 - (d) Центральность по посредничеству
 - (e) PageRank
4. Формулировка выводов и описание рекомендаций

Исследование ключевых показателей:

Степенная центральность:

В этом показателе заложена идея о том, что самая важная вершина - та, что имеет наибольшее число связей с другими вершинами. Так как у нас не взвешенный и не ориентированный граф, нам достаточно посчитать количество инцидентных ребер, чтобы найти степени для каждой вершины. Сделаем это, написав код на Python (реализация с комментариями - в приложенном файле).

Рассмотрим график распределения степеней, визуализировав найденные значения:



Видно, что наибольшее количество социальных взаимодействий у дельфина Grin. Значит, при данном подсчете он самый влиятельный в этой сети.

Для него получилось 12 инцидентных ребер.

Также можем отметить, что:

1) В популяции нет отщепенцев, которые ни с кем активно не взаимодействуют (в этом можно убедиться, сложив все координаты у для каждого целого значения x в графике распределения, сумма будет равняться 62 - размеру популяции).

2) Минимальное количество частых социальных связей для дельфина этой популяции равно 1.

3) Среднее количество активновзаимодействующих партнеров для дельфина в этой популяции - 5.

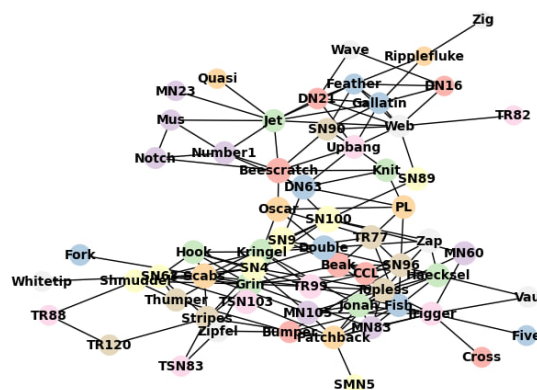
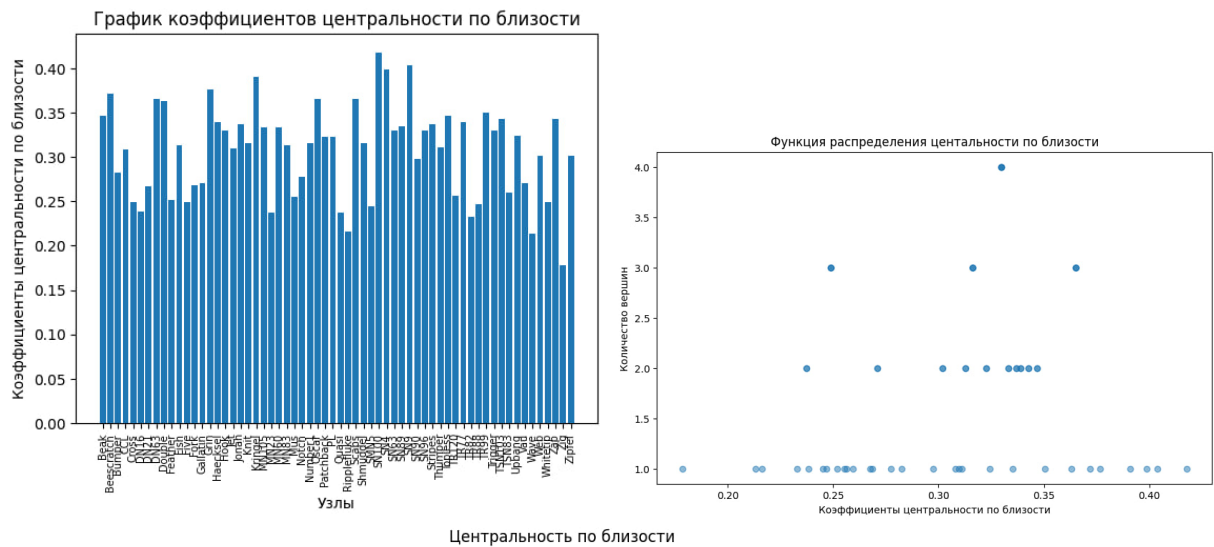
Центральность по близости:

Идея этого показателя в том, что наиболее влиятельной является вершина, находящаяся ближе всех к остальным участникам сети.

Для каждого узла коэффициенты рассчитываются по данной формуле:

$$C_i = \frac{1}{\sum_j d_{ij}}$$

Участник с наибольшим коэффициентом будет считаться самым влиятельным в модели.



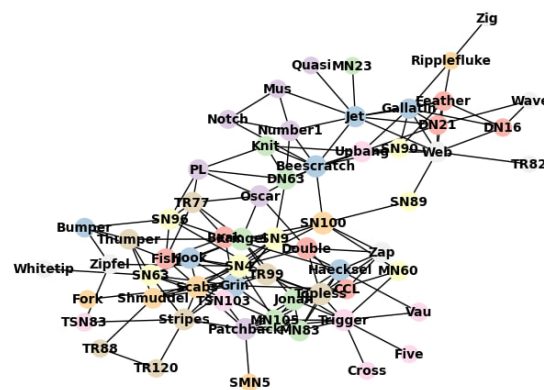
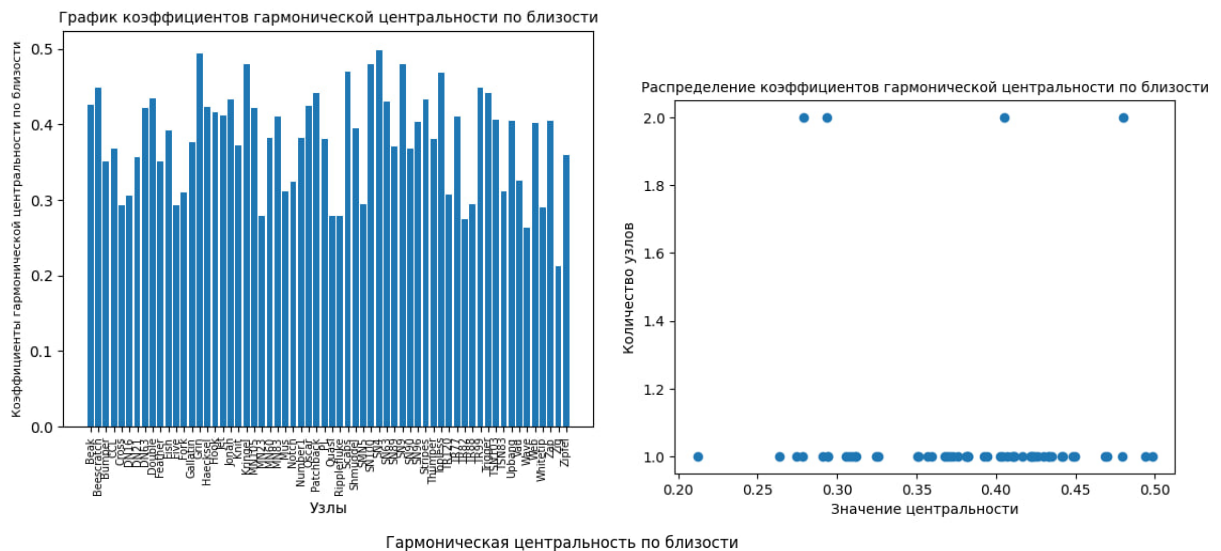
Какие выводы здесь можно сделать:

- 1) Значение всех коэффициентов лежит в диапазоне $[0.178; 0.418]$ – это достаточно небольшой диапазон, меньше половины от максимального возможного.
- 2) Соответственно, наиболее важный дельфин в этой сети имеет коэффициент 0.418, это SN100. Наименее влиятельный - дельфин Zig (0.178).

Гармоническая центральность по близости:

Идея та же самая, что и в центральности по близости, только в этот раз коэффициенты каждого узла нужно рассчитать по другой формуле:

$$C_i = \sum_j \frac{1}{d_{ij}}$$



Теперь самый влиятельный дельфин - это SN4.

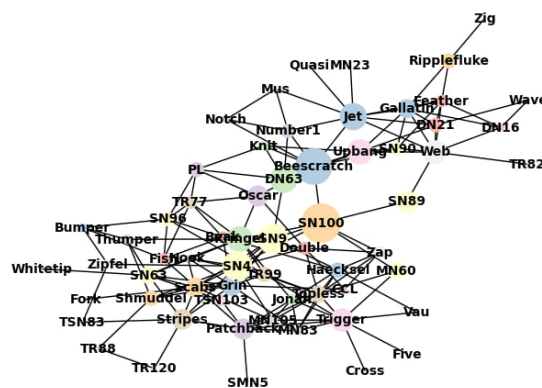
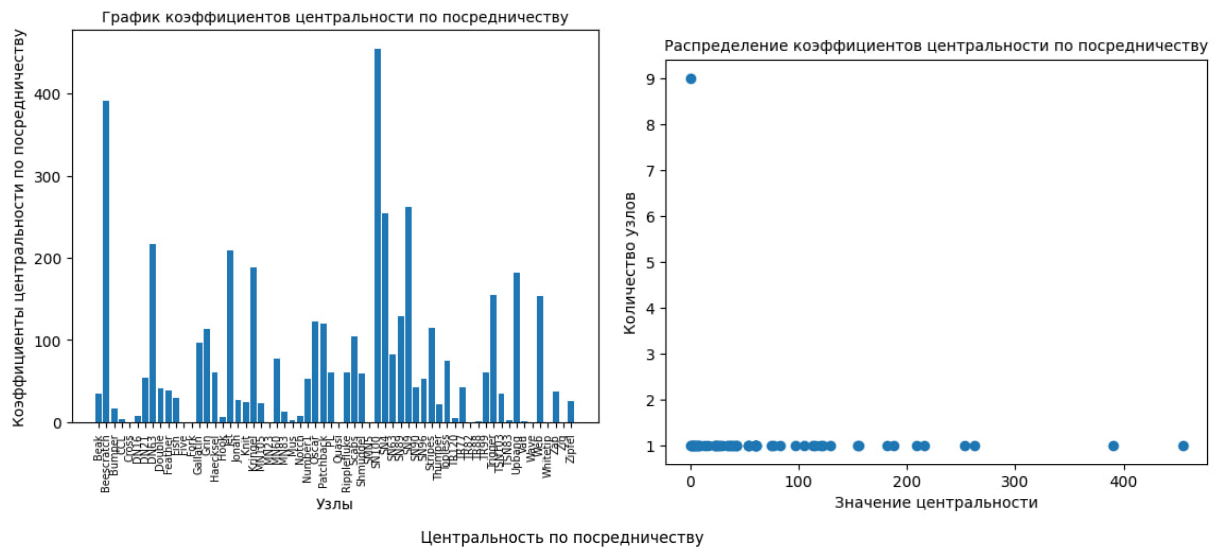
Такой подсчет дает больше шансов получить высшую оценку не тем, кто имеет много общительных соседей, а тем, кто сам имеет много связей с особями, имеющими среднее количество связей. (так как учитывается конкретное значение длины пути, а не совокупность их длин).

Центральность по посредничеству:

При данном способе подсчета наиболее значимым элементом является тот, через который проходит наибольшее количество кратчайших путей.

Коэффициенты центральности по посредничеству для каждой вершины считаются следующим образом:

$$b_i = \sum_{u,v \in V} \frac{\sigma_{u,v}(i)}{\sigma_{u,v}}$$



Получили, что наиболее важный дельфин в этой сети - SN100. Его показатель равен 454.274. Наименее важные дельфины при таком случае подсчета - Zig, Cross, Five, Fork, MN23, Quasi, SMN5, TR82, Whitetip - через них не проходит ни один кратчайший путь, соответственно их коэффициенты равны 0.

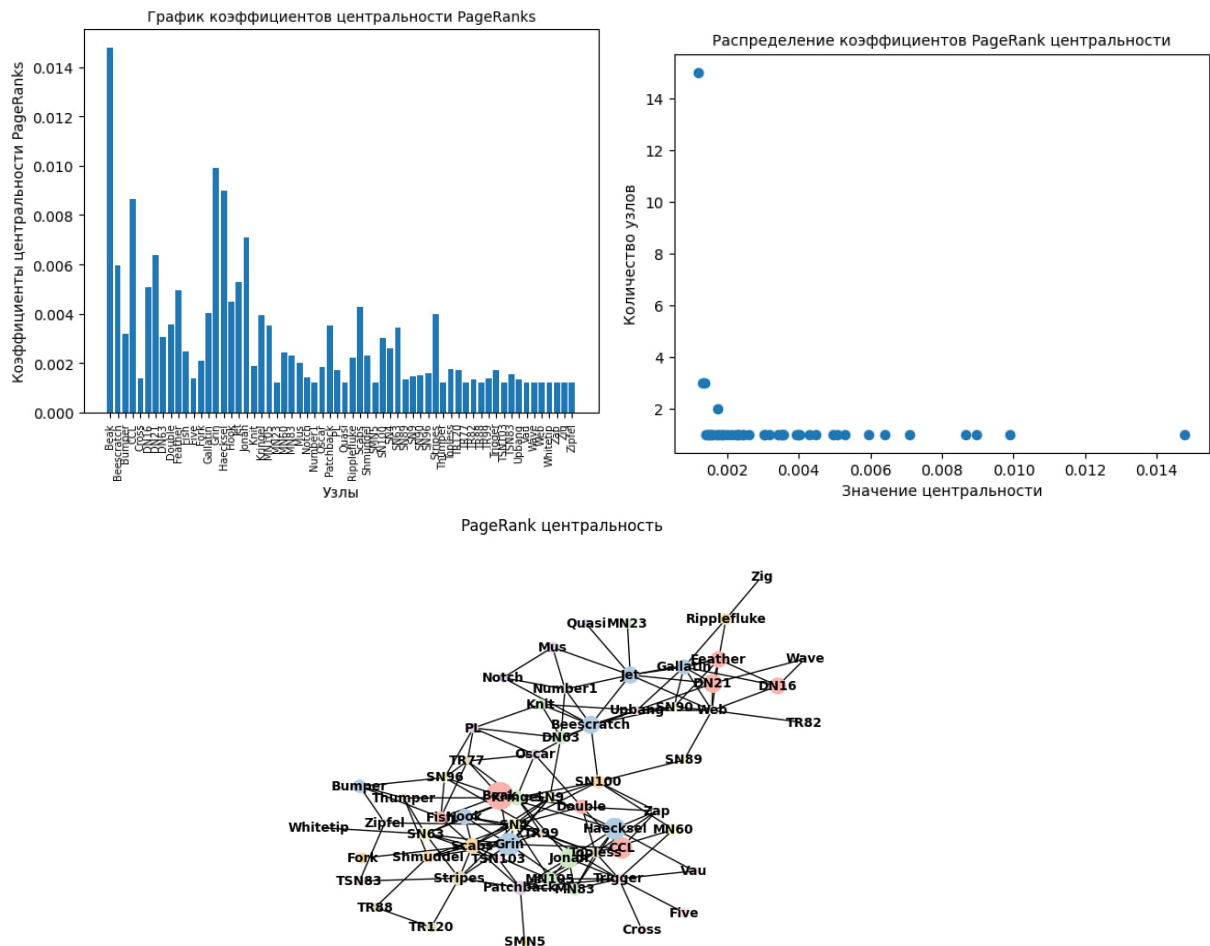
Теперь давайте разберемся, почему SN100 самый важный в этой модели?

Во-первых, количество кратчайших путей, проходящих через узел, связано со степенью близости рассматриваемой вершины к остальным, так как существует высокая вероятность, что какой-либо кратчайший путь пройдет через ближайший узел. SN100 является важным элементом как в центральности по близости (топ-1), так и в гармонической центральности по близости (топ-3), поэтому для такого результата уже была предпосылка.

Во-вторых, у данной вершины много соседей, имеющих высокие коэффициенты по этой центральности, (Beescratch, SN9, SN4 - входят в топ 5) а значит, эта вершина с большой вероятностью может стать продолжением какого-то кратчайшего пути.

PageRanks:

В оригинале этот показатель используется для вычисления вероятности того, что пользователь сети зайдет на тот или иной сайт. Хотя наша сеть от интернета и отличается порядком, мы все равно можем изучать благодаря этой сети некоторые задачи, например, до какого дельфина дойдет с большей вероятностью та или иная информация.



Из распределения видно, что вероятность, что до какого-то дельфина дойдет информация, которую передают между собой две другие особи в стае, достаточно мала.

Нетрудно заметить, что разница между соседними вершинами в рейтинге незначительная (либо нулевая, либо 0,001). Показатели первых 20 дельфинов, среди которых есть все лидеры по другим показателям, отличаются не более чем на 0.012 - очень малое число.

Результаты:

1) Если нам нужно изучить взаимодействие дельфинов в малых группах с разными вводными, стоит разбивать их на группы по 6 дельфинов (степенная центральность). Это даст наиболее чистый результат, из которого можно понять влияние изучаемого фактора, так как вероятность получения информации любым дельфином в сети отличается незначительно от особи к особи (центральность PageRank).

2) Если нас интересуют различия в поведении во время коммуникации самых социально-активных дельфинов и наименее социально-активных особей, можно рассмотреть дельфинов с наибольшей степенной центральностью (например, Grin, SN4, Topless, Scabs, Trigger) и с наименьшей степенной центральностью (например, Zig, TR82, Quasi, MN23, SMN5).

3) Если мы хотим исследовать как изменяется какая-либо информация во время передачи по сети или любую подобную задачу, стоит брать промежуточные данные дельфина SN100, так как он лучший по центральности по посредничеству, а значит, через него проходит максимальное количество путей, связывающих две какие-то особи.

4) Если мы хотим исследовать сеть на скорость распространения какой-либо информации по сети или любую подобную задачу, стоит обратить внимание на среднее достижение информации до SN100 и SN4, так как они находятся потенциально ближе всего в области

достижимости каждой вершины. С рандомным узлом проводить это имеет меньший смысл, так как в целом коэффициенты центральностей по близости малы.