Dynamic and thermodynamic models of adaptation

## Dynamic and thermodynamic models of adaptation

## Abstract

### Key words

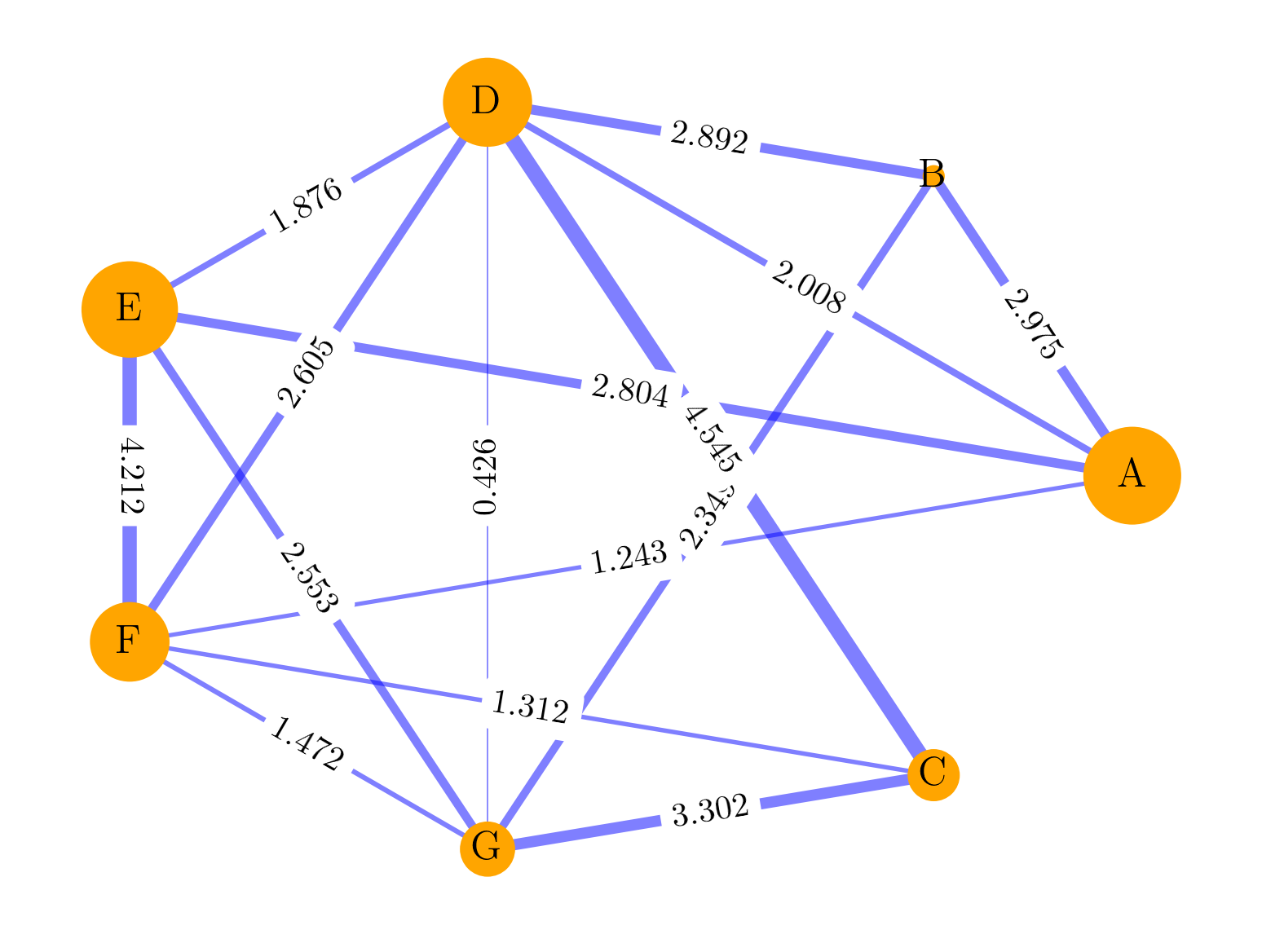
## Introduction

Теория графов позволила по-другому взглянуть на живые организмы и протекание биологических процессов. В отличие от простых моделей, описывающих конкретные процессы, сетевой анализ позволяет учесть множество взаимодействующих факторов, которые можно представить в качестве узлов графа, важность этих факторов в различных процессах (потенциальный атрибут веса узлов).

Факторы неким образом взаимодействуют между собой и характер их взаимодействия возможно неким образом оценить. Такая оценка может быть представлена, как атрибут или вес ребра между факторами-узлами.

Безусловно, в биологических процессах и живых системах характеристики протекающих процессов крайне изменчивы и имеют сложные распределения. Таким образом, представляется крайне сложным определить некие модели, описывающие взаимодействия между характеристиками и оценивающие вес самих факторов-узлов. Такие модели должны быть крайне сложными и нелинейными. Выбор конкретных моделей может быть осуществлен эмпирически.

Более того, сетевой анализ позволяет еще и динамически оценивать системы, как изменения во времени структуры графа и его характеристик, что еще более приближает данный вид анализа к реалистичному описанию биологических систем. При оценке динамических колебаний в таких структурах перспективными напрвавлениями являются оценка реакций систем на какие-либо внешние раздражители, оценка адаптации, кризисных ситуаций и полной дезадаптации систем.

 **Figure 1.** Пример взвешенного графа. Вершины и ребра имеют размер и толщину в зависимости от их весов. Граф может быть и направленным, что может сделать модель более реалистичной.

Наиболее важным есть и остается вопрос о выборе наиболее репрезентативной описательной модели взаимосвязей в графе.

В таком случае, принципиальным вопросом является выбор той системы, которую мы рассматриваем в виде сети.

Такими системами могут быть:

1. Некие индивидуальные экспериментальные единицы одного множества (пациенты, колонии бактерий, лабораторные животные, белковые системы и т.д.) - в таком случае, граф и будет представлять такую единицу. Все остальные единицы должны иметь сходный набор факторов, различия же будут отмечаться лишь по весам вершин и ребер;
2. Одна единственная система, динамически меняющаяся во времени. В таком случае, ее состояния в различных временных точках могут быть представлены в виде графов. Такие сети также должны иметь сходный наборр характеристик (узлы), чьи веса и взаимные связи (атрибуты ребер) меняются со временем;
3. Некие индивидуальные экспериментальные единицы одного множества, свойства которого меняются со временем. Таким образом, будет получен набор графов, характеризующий каждую индивидуальную экспериментальную единицу, как в случае (1), а также для каждой такой единицы с учетом временных точек возможно будет построить динамические графы, как в случае (2).

Каким же образом оценивать связи и веса в таких графах?

Если говорить о моделях, работающих на уровне биохимических показателей, иммунологических реакций, описывающих генетические полдиморфизмы и т.д. - подавляющее большинство характеристик экспериментальной единицы являются количественными. При рассмотрении пациента в рамках экспериментальной единицы, его истори заболевания и результатов обследований, других характеристик, большая часть из них являются качественными.

В настоящей главе авторы приводят протокол для моделей, описывающих взаимодействие количественных переменных. Мы рассмотрим методы построения наиболее изученных моделей, которые уже неоднократно были использованы в экспериментальных исследованиях:

* паренклитические графы с использованием порога[1];
* взвешенные паренклитические графы[2], [3];
* взвешенные синолитические графы[4];
* корреляционные графы с использованием порога[5];
* взвешенные корреляционные графы[5].

Существуют несколько логических этапов, связанных с рассмотрением экспериментальных единиц в разрезе сетевого анализа:

1. Выбор задач по работе с системой, которую мы хотим представить в виде сети - будет ли это задача классификации или оценка динамических изменений?
2. Предварительная трансформация данных;
3. Определение характеристик, которые будут являться узлами будущего графа;
4. Выбор метода, который позволит оценить вес ребер (фактически, отражает выбор модели из списка выше);
5. Построение графа с использованием порога или взвешенного графа;

Далее мы можем пойти одним из двух путей:

* + Расчет ряда показателей для каждого графа и последующая работа с этими показателями, как с новыми характеристиками экспериментальной единицы. Мы затем можем трансформировать данные и отправить их на обучение некоего классификатора (алгоритм машинного обучения, нейронная сеть и т.д.);
  + Отправка целого графа или массива из графов для обучения в графовую нейронную сеть (GNN).

## Materials

### Взвешенные паренклитические сети (wLRPA)

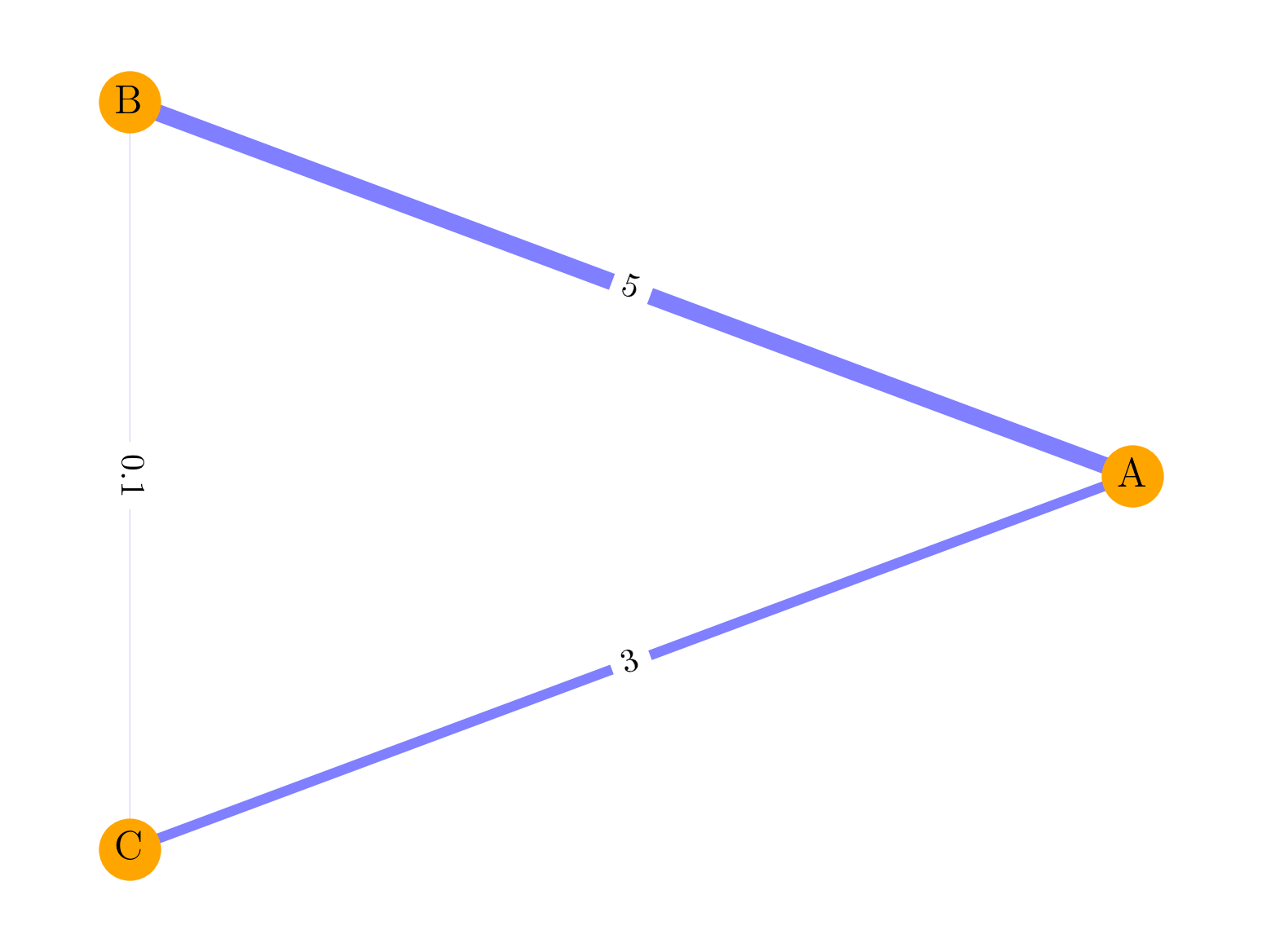
1. Выделяется экспериментальные единицы, которые являются контрольной группой - данная группа представляет множество “здоровых” или “нормальных” единиц.
2. Выделяются попарные комбинации без повторений между характеристиками единицы. Например, при наличии 3-х характеристик единицы [А, В, С] попарные комбинации будут следующие: [A-B, A-C, B-C]. Для дальнейшего построениия графов и работы с ними важно, чтобы сами комбинации и порядок элементов были идентичными для всех экспериментальных единиц.
3. Для каждой попарной комбинации характеристик (, ) всей контрольной группы строится уравнение линейной регрессии:

* , где
  + - экспериментальная единица,
  + - остаток в уравнении линейной регрессии
* Таким образом, для каждой пары характеристик (, ) имеется массив с отклонениями [], где - экспериментальная единица.

1. Массивы с отклонениями преобразуются в абсолютные Z-значения, |Z-scores|:

* , где
  + - экспериментальная единица,
  + - среднее отклонений из уравнения линейной регрессии для характеристик (, ) по контрольной группе,
  + - стандартное отклонение значений отклонений по контрольной группе.

1. Полученные абсолютные Z-scores и становятся весами ребра между узлами (, ).
2. После получения весов для каждой пары узлов по каждой экспериментальной единице строится взвешенный граф , где V - множество узлов графа, соответствующих характеристикам экспериментальной единицы, E - множество ребер графа, соответствующих полученным абсолютным Z-scores.

*  **Figure 1.** Пример взвешенного графа гипотетической экспериментальной единицы паренклитической сети. Толщина ребер соответствует величине веса абсолютного Z-значения между характеристиками A, B и C.
  + итогом является изменение представления структуры данных по каждой экспериментальной единицы, которая теперь предстает перед нами в виде сети характеристик, ассоциированных между собой.

1. Для полученного взвешенного графа могут быть рассчитаны ряд его характеристик, среди которых рекомендованными являются:
   * описательные статистики (количество нулей, минимум, максимум, среднее, стандартное отклонение и любые другие) для таких показателей графа, как closeness, betweenness, edge betweenness, page rank, eigen centrality, authority score, node degree, edge weights;
   * the vector of all nodes degrees in a graph.
2. Таким образом, осуществляется изменение и представления, и размерности уже сетевой структуры, и мы вновь имеем дело уже с табличными рядами данных, характеризующих экспериментальную единицу.
3. Все предшествующие манипуляции проводились для контрольной группы, куда входили некие “здоровые” или “нормальные” единицы. При появлении некоей новой экспериментальной единицы или нескольких экспериментальных единиц проводимые манипуляции следующие. Для каждой новой экспериментальной единицы для каждой пары характеристик (, ) строится уравнение линейной регрессии, но используются коэффициенты и , которые были получены для контрольной группы. После этого оценивается величина , где - новая экспериментальная единица.
4. Полученные массивы отклонений для новых экспериментальных единиц нормализуются относительно среднего и стандартного отклонения для контрольной группы:

* , где
  + - новая экспериментальная единица,
  + - среднее отклонений для характеристик (, ) по контрольной группе,
  + - стандартное отклонение значений отклонений по контрольной группе.
* В результате, получены абсолютные Z-scores для новых единиц.

1. Впоследствии для новых единиц проводятся манипуляции 5. - 7. для создания графов и вычисления их характеристик.

### Взвешенные паренклитические сети (wKDEPA)

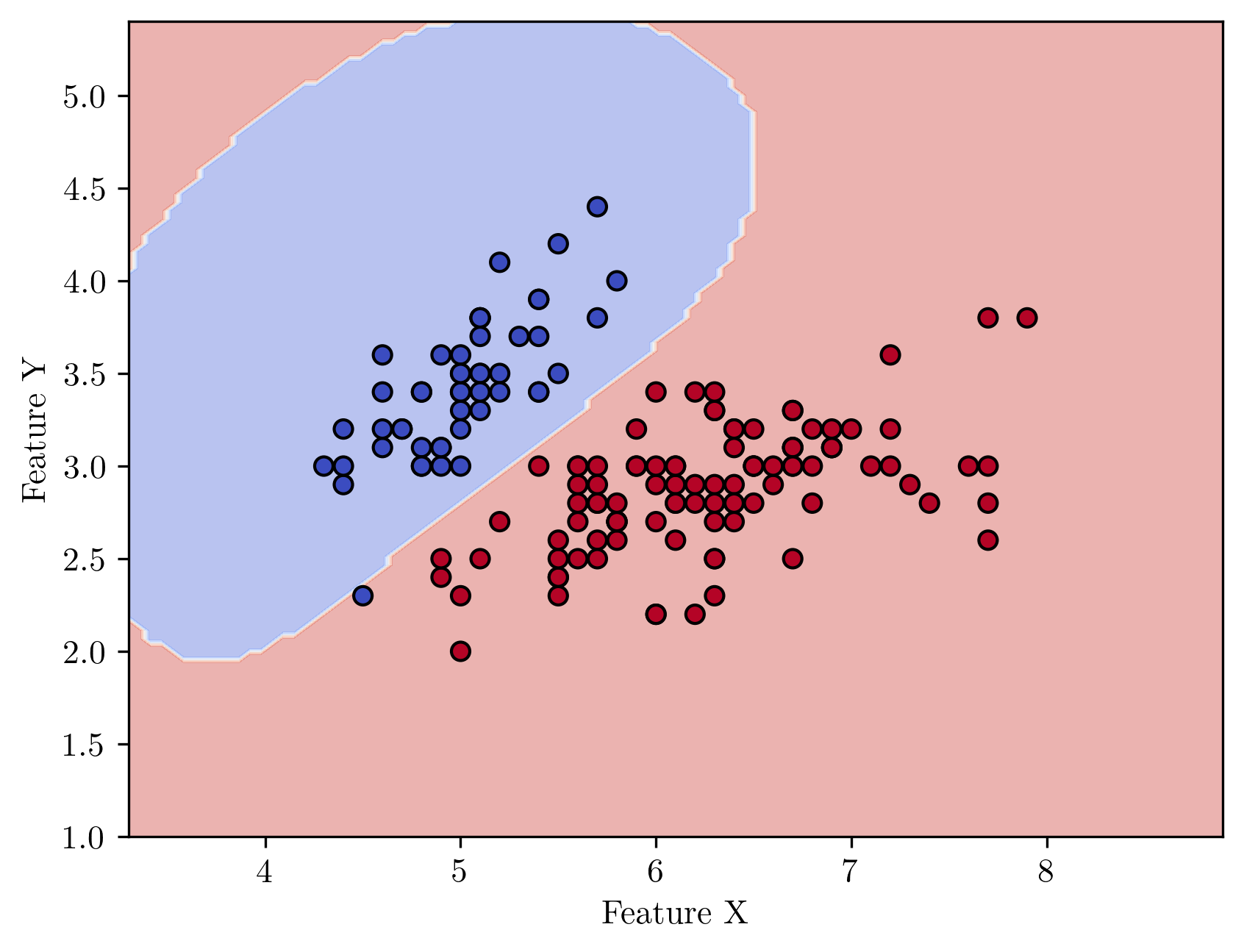
1. Снова выделяются экспериментальные единицы, которые являются контрольной группой или группой “нормальных” единиц. Также выделяются попарные комбинации характеристик, однако уже с целью построения two dimensional kernel density estimation.
2. Для каждой комбинации характеристик (, ) создается квадратная two dimensional kernel density matrix, получившая название контурной матрицы. *Количество строк и столбцов матрицы минимально определяется длиной векторов и , при этом может быть увеличено. (Почему?)*. В зависимости от параметров функции (например, с помощью функции kde2d в R или в Python с помощью функций mgrid из модуля numpy, а затем с расчетом контурной матрицы с помощью функции stats.gaussian\_kde модуля scipy), создающей контурную матрицу, количество рядом и столбцов такой матрицы может варьировать.
3. Так как полученная матрица имеет …

### Синолитические сети (wSA)

В основе синолитического подхода лежит выстраивание границы между классами с помощью того или иного классификатора. В качестве примера в литературе был рассмотрен метод опорных векторов (SVM) with the radial basis function kernel, а также логистическая регрессия [2]. Однако может быть использован любой подходящий классификатор.

Последовательность манипуляций при использовании синолитического подхода следующая:

1. Используются экспериментальные единицы, относящиеся к классам “нормальных” и “аномальных” (“здоровых” и “больных”), а не только контрольная группа целиком из “нормальных” экспериментальных единиц.
2. Вновь определяются попарные комбинации характеристик. В качестве примера классификатора используем SVM with the radial basis function kernel, однако возможно использовать любой классификтор.
3. Для каждой комбинации характеристик (, ) с помощью классификатора определяется граница, разделяющая классы. Иными словами, для каждой комбинации создается отдельный классификатор, в попытке провести границу между классами. Визуально разделение может быть представлено следующим образом.

 **Figure 2.** Разделение двух классов (красные и синие) с помощью radial SVM. В связи с тем, что при синолитическом подходе используются единицы разных классов.

1. С помощью классификатора определяется вероятность принадлежности каждой точки на плоскости (, ) к тому или иному классу. Т.к. каждая точка обозначает отдельную экспериментальную единицу, полученные вероятности и становятся весами ребер между узлами, отражающими характеристики (, ). Таким образом, каждая экспериментальная единица получает вес ребра между узлами (, ).
2. В результате, все экспериментальные единицы меняют свое представление на взвешенные графы.

### Корреляционные графы

Корреляционные графы наиболее часто используются при возможности получить несколько последовательных измерений одних и тех же характеристик у одной и той же экспериментальной единицы. Если у нас есть множество таких экспериментальных единиц, у нас должна быть возможность проводить одно ито же количество таких последовательных измерений, а характеристики должны быть идентичными.

В основе такой концепции лежит гипотеза Александра Горбаня [5], согласно которой в некоей системе с множеством характеристик при условии стабильности имеются некие базисные корреляции и оказатели дисперсии между характеристиками .

При условии возникновения некоего “стресса” в системе возрастает скоррелированность характеристик, а также дисперсия .

В дальнейшем, система либо снова возвращается к стабильности (при этом мы будем наблюдать возвращение показателей корреляции и дисперсии до базисных значений , либо наступит этап дезадаптации (или кризиса), при котором корреляции значительно снижаются (в том числе ниже уровня базисных), а дисперсия остается большой или возрастает еще больше .

Общая инструкция для построения корреляционного графа следующая.

1. Определяем характеристики экспериментальной единицы, между которыми мы будем попарно определять корреляции, например, . По каждой из этих характеристик мы имеем набор значений во времени, временные точки должны совпадать.
2. Исследователи выбирают начальное временное “окно,” например, измерения по характеристикам на момент времени и некий эмпирический размер этого окна. Предположим, величина окна выбрана в 20 временных точек. Тогда мы получаем 19 предшествующих временных точек по каждой характеристике и измерения на момент времени (т.е., всего 20 измерений). Соответственно, для характеристик мы получаем по 20 временных значений с учетом для каждой экспериментальной единицы.

| Временная точка | A | B | C |
| --- | --- | --- | --- |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |
| … | … | … | … |
|  |  |  |  |

1. Названия характеристик становятся узлами графа. Между каждыми 2-мя характеристиками определяется корреляция и величина коэффициента корреляции становится атрибутом ребра между узлами.
2. Метод оценки корреляции напрямую связан с типами подлежащих распределений характеристик. В целом, для создания корреляционных графов могут быть использованы метод по Пирсону, по Спирману, Cramer’s V, Tau и другие. отдельно стоит уделить внимание значимости корреляционной связи. При отсутствии значимой корреляции логичным представляется не строить ребро между соответствующими узлами, в то время как наличие значимой корреляции позволяет построить ребро и присвоить ему вес в виде корреляционного коэффициента.
3. Таким образом, мы получили корреляционный граф на момент времени .
4. Следующий этап - построение аналогичного графа для следующего момента времени : временное окно из 20 временных точек переезжает на один временный промежуток вперед и повторяется полная последовательность 3. - 5. Создается новый граф для момента времени и так далее.

## Methods

## Notes

## References

1. Zanin M, Boccaletti S. Complex networks analysis of obstructive nephropathy data. *Chaos: An Interdisciplinary Journal of Nonlinear Science*. 2011;21(3):033103. doi:[10.1063/1.3608126](https://doi.org/10.1063/1.3608126)

2. Nazarenko T, Whitwell HJ, Blyuss O, Zaikin A. Parenclitic and Synolytic Networks Revisited. *Front Genet*. 2021;12. doi:[10.3389/fgene.2021.733783](https://doi.org/10.3389/fgene.2021.733783)

3. Whitwell HJ, Blyuss O, Menon U, Timms JF, Zaikin A. Parenclitic networks for predicting ovarian cancer. *Oncotarget*. 2018;9(32):22717-22726. doi:[10.18632/oncotarget.25216](https://doi.org/10.18632/oncotarget.25216)

4. Krivonosov M, Nazarenko T, Bacalini MG, et al. Age-related trajectories of DNA methylation network markers: a parenclitic network approach to a family-based cohort of patients with Down Syndrome. *bioRxiv*. Published online January 2022:2020.03.10.986505. <https://doi.org/10.1101/2020.03.10.986505>

5. Gorban AN, Smirnova EV, Tyukina TA. Correlations, Risk and Crisis: From Physiology to Finance. *arXiv*. Published online May 2009. doi:[10.1016/j.physa.2010.03.035](https://doi.org/10.1016/j.physa.2010.03.035)