**Разработка программного модуля для нахождения тройных взаимодействий в биологических сетях с учётом временных задержек**

**Введение**

Анализ неструктурированной информации берет своё начало в 1958 году в работе [REFERENCE]. Однако, только в начале 21 века развитие технологий позволило восполнить исследовательский интерес, в таких работах как [REFERENCE]. Одной из областей анализа неструктурированной информации является восстановление структуры биологических сетей. К ним относят генные сети(ответственные за протекание физических процессов в организме[REFERENCE]), биологические нейронные сети(ответственные за распространение сигнала в мозге[REFERENCE]), биополимеры(элементы, участвующие в биологических процессах[REFERENCE]). Изучение каждой из этих сетей позволяет лучше узнать физиологические процессы, протекающие в организме. Так, например, в 2012 году в работе[REFERENCE], группа учёных смогла выявить генную сеть, ответственную за развитие шизофрении, и нашла взаимосвязь этой сети с сетью, ответственной за аутизм. Данной открытие позволяет лучше понять причинно-следственную характеристику развития такой болезни как шизофрения.

Алгоритмы, восстанавливающие структуру сетей строятся на анализе взаимодействий между элементами сети. Поэтому одним из главных этапов нахождения структуры биологических сетей является выявление как парных взаимодействий[REFERENCE], так и взаимодействий более высокого порядка [REFERENCE](согласованного взаимодействия более чем двух элементов). Взаимодействия высокого порядка встречаются в реальных биологических системах. Так, например, при воссоздании структуры белка, должны использоваться методы, учитывающие тройные взаимодействия, поскольку белки являются компактной структурой, в которой не обойтись описанием только парных взаимодействий[REFERENCE]. Также, в более ранних трудах было показано[REFERENCE], что алгоритмы, учитывающие взаимодействие порядка более двух, дают более точные результаты структуры сети. Поскольку учёт взаимодействий высокого порядка является вычислимо трудоемкой задачей[REFERENCE], на текущий момент ограничиваются учетом взаимодействий третьего порядка. Взаимодействий порядка более трёх учитывают с системах с небольшим количеством элементов(порядка 10-100). Особый интерес представляет нахождение взаимодействий с учётом временных задержек, так как в реальных биологических процессах, реакция того или иного компонента сети может проявляться с задержкой[REFERENCE]. На текущий момент, существуют алгоритмы, учитывающие взаимодействия третьего порядка[REFERENCE], а также алгоритмы, учитывающие парные взаимодействия с временными задержками.

В данной работе предлагается алгоритм, учитывающий взаимодействия третьего порядка с временными задержками. За основу был принят алгоритм MIDER[REFERENCE], который умеет выявлять парные взаимодействия с учетом временных задержек. Данный алгоритм был доработан учетом тройных взаимодействий, а также был изменён алгоритм учёта временных задержек[REFERENCE]. Были предложены методы оптимизации задачи нахождения тройных взаимодействий. Показаны результаты анализа реальной биологической сети, предложена реализация алгоритма на MatLab, доступная по ссылке [REFERENCE].

**Заключение**

В данной работы был создан программный комплекс для нахождения тройных взаимодействий в биологических сетях с учётом временных задержек. На текущий момент, этот алгоритм является первым, учитывающим тройные взаимодействия, а также временные задержки.

За основу был взят алгоритм MIDER, его доработка учетом тройных взаимодействий позволила улучшить его точность, что было продемонстрировано на площадке DREAM Challenge, на испытаниях [REFERENCE]. На каждом из испытаний, продемонстрированная точность была минимум на [INSERT] выше, чем алгоритм без учёта тройных взаимодействий. Применение алгоритма DTW для учёта временных задержек также позволило улучшить точность. Данный алгоритм может быть эффективно применён как для исследования реальных биологических сетей, так и для других задач анализа взаимодействий, на выборках объёмом не более [INSERT] элементов. Предложенные методы оптимизации, такие как регуляризация Шмидта, выделение значимого подпространства, позволяют уменьшить вычислительную сложность алгоритма минимум в [INSERT] раза, и дальнейшие изыскания в области оптимизации алгоритма позволят увеличить возможное количество элементов сети до порядка [INSERT].

Нахождение взаимодействий более высокого порядка(четырёх и более) является перспективной задачей, поскольку учёт таких взаимодействий позволяет получить более точную картину сети. В особенности, если учитываются временные задержки. Дальнейшая работа в этой области будет направлена на увеличение порядка учитываемых взаимодействий, а также на оптимизацию существующих методов.