

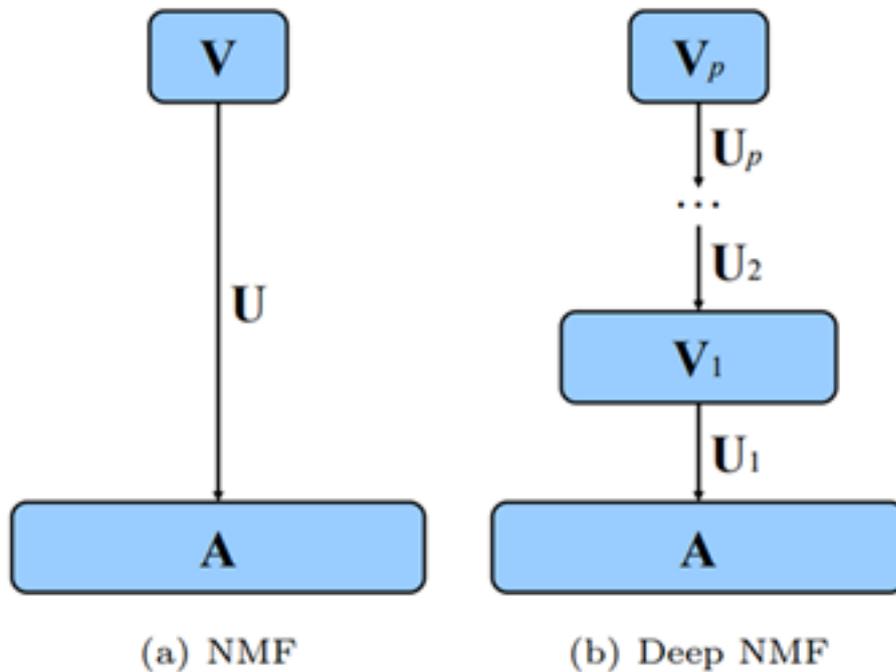
## ИССЛЕДОВАНИЕ МОДЕЛИ DEEP AUTOENCODER-LIKE NON-NEGATIVE MATRIX FACTORIZATION

**Савельев Павел Николаевич**

магистрант, Самарский национальный исследовательский университет им. академика С.П. Королева, РФ, г. Самара

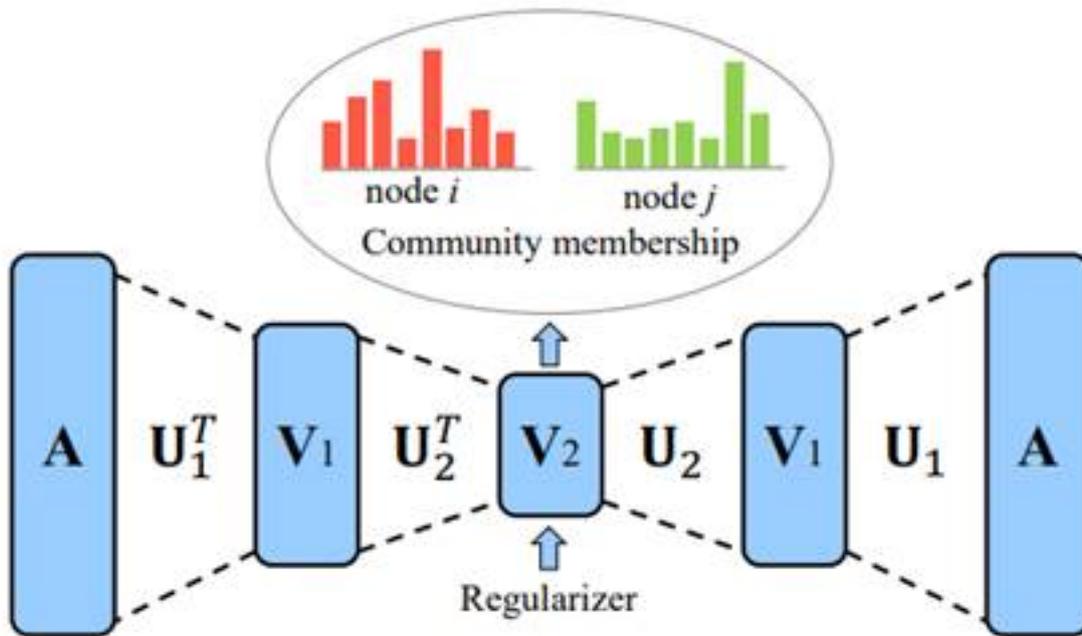
Структуры сообществ повсеместны в реальных комплексных сетях. Задача обнаружения сообщества по этим сетям имеет первостепенное значение в различных сферах. В последнее время факторизация неотрицательной матрицы (NMF) получила широкое распространение для обнаружения сообществ из-за её большой интерпретируемости и естественной пригодности для определения членства узла в сообществе. Однако существующие подходы обнаружения сообщества на основе NMF являются поверхностными. Они ищут членство в сообществе, напрямую сопоставляя исходную сеть с областью членства в сообществе. Учитывая сложные и разнообразные структуры топологии реальных сетей, весьма вероятно, что отображение между исходной сетью и пространством членства в сообществе содержит довольно сложную иерархическую информацию, которая не может быть интерпретирована классическими подходами на основе поверхностного NMF. Совсем недавно появилась новая модель, называемая Deep Autoencoder-like NMF (DANMF), для обнаружения сообществ. Подобно глубокому автоэнкодеру, DANMF состоит из компонента кодера и компонента декодера. Эта архитектура позволяет DANMF изучать иерархические отображения между исходной сетью и окончательным назначением сообщества с неявными скрытыми атрибутами низкого и высокого уровня исходной сети, изученными на промежуточных уровнях.

Глубокий автоэнкодер является отличной схемой для сокращения разрыва между абстракцией более низкого уровня и абстракцией более высокого уровня исходных данных. Руководствуясь глубоким автоэнкодером, можно утверждать, что, дополнительно разлагая отображение  $U$  таким образом, что каждый фактор добавляет дополнительный уровень абстракции подобия между узлами от более низкого уровня к более высокому уровню, можно затем получить лучшее сходство на уровне сообщества между узлами (т. е. более точная матрица членства в сообществе  $V$ ), как показано на рисунке 1. Например, можно узнать сходство между узлами из близости первого порядка, степени ассортативности, структурной идентичности и, наконец, сходства на уровне сообщества.



**Рисунок 1. (a) Архитектура NMF. (b) Архитектура глубокой NMF. Deep NMF изучает иерархию скрытых атрибутов, которые помогают раскрыть окончательное членство узлов в сообществах**

Подобно глубокому автоэнкодеру, компонент кодера пытается преобразовать исходную сеть в пространство членств в сообществах с неявными скрытыми низкоразмерными атрибутами, полученными на промежуточных уровнях. Каждый промежуточный уровень интерпретирует сходство между узлами на разных уровнях детализации. Компонент декодера является симметричным с компонентом кодера. Он стремится восстановить исходную сеть из пространства членств в сообществах с помощью иерархических отображений, полученных в компоненте кодера. В отличие от традиционных методов обнаружения сообщества на основе NMF, которые учитывают только функцию потерь компонента декодера, DANMF объединяет и компонент кодера, и компонент декодера в единую функцию потерь. Таким образом, DANMF наследует способность к обучению глубокого автоэнкодера, в то же время улучшая интерпретируемость модели благодаря неотрицательным ограничениям, и он подходит как для обнаружения непересекающихся сообществ, так и для обнаружения пересекающихся сообществ. Кроме того, DANMF включает в себя регуляризатор графа для соблюдения внутренней геометрической структуры пар узлов. Общая структура DANMF показана на рисунке 2. Для иллюстрации глубина установлена на 2. Компонент кодера (левая часть) преобразует сеть в пространство членства в сообществе. Компонент декодера (правая часть) восстанавливает сеть из пространства членства в сообществе.

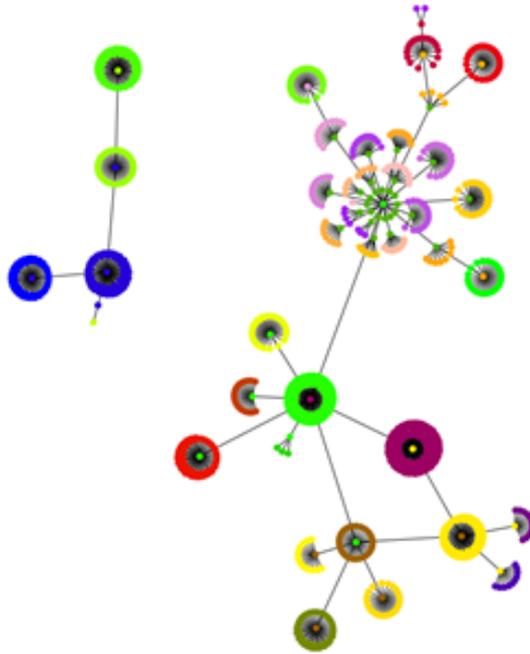


*Рисунок 2. Архитектура DANMF*

Рассмотрим работу DANMF на наборе данных, состоящем из 1005 узлов и 25571 рёбер, а также сравним его результаты с классическими методами Community Detection и Connected Components.

Набор данных был создан с использованием данных электронной почты из большого европейского исследовательского учреждения. В нём содержится анонимная информация обо всех входящих и исходящих сообщениях электронной почты между членами исследовательского учреждения. В сети есть ребро  $(u, v)$ , если человек отправил человеку по крайней мере одно электронное письмо. Электронные письма представляют только связь между членами организации, а набор данных не содержит входящие сообщения или исходящие сообщения.

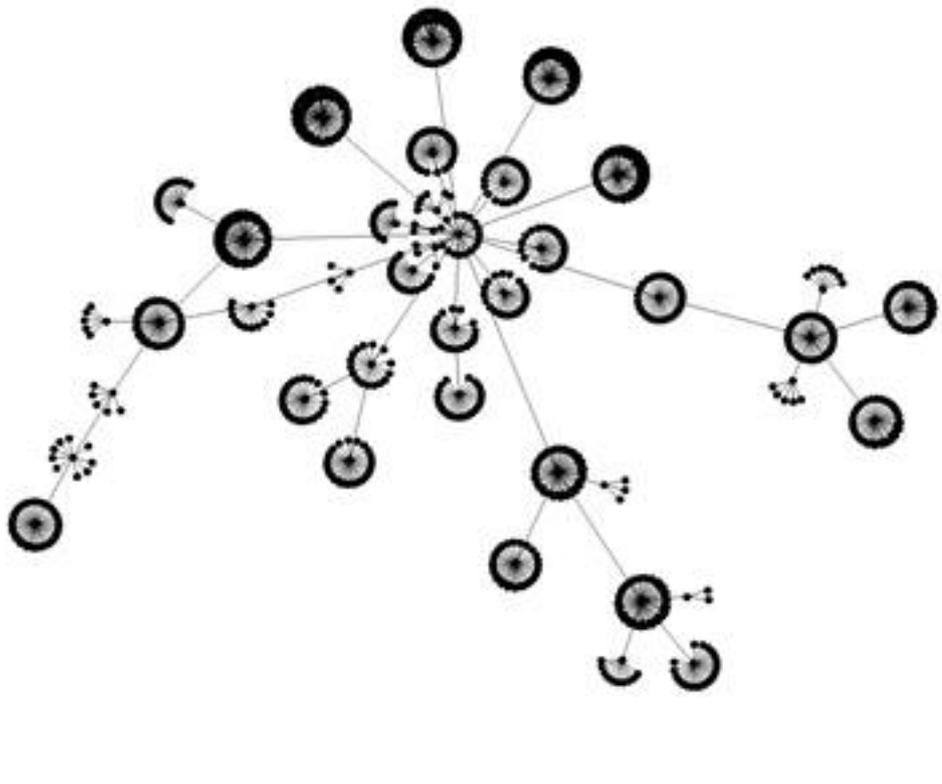
Идеальное распределение на группы, заданное в начальном наборе данных представлено на рисунке 3.



*Рисунок 3. Идеальное распределение на сообщества*

Для работы DANMF количество слоев модели было установлено 1005-256-128-42, исходя из того, что количество сообществ в наборе данных равно 42.

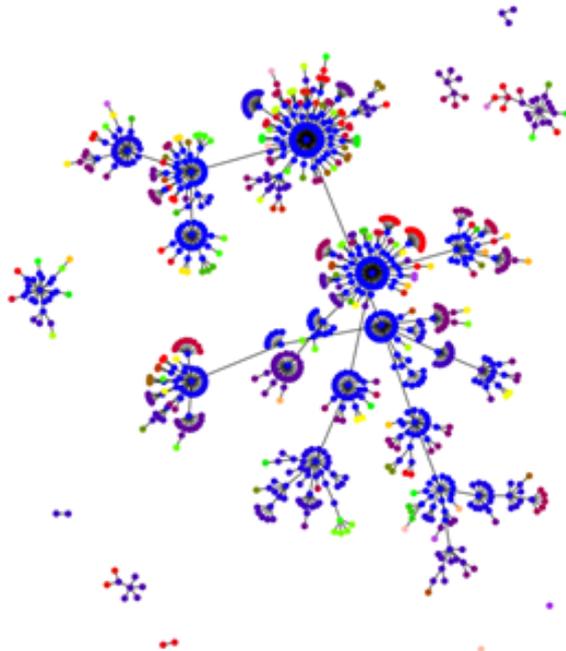
На рисунке 4 изображен результат работы при вышеуказанной конфигурации слоев и методе преобучения shallow.



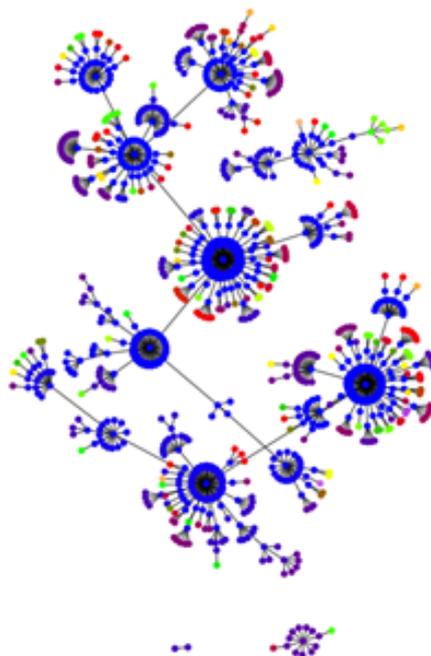
*Рисунок 4. Результат работы DANMF, метод shallow*

Как можно заметить, результат довольно похож на исходную картину.

Сравним теперь результаты с методами Community Detection и Connected Components. Результаты работы алгоритмов Community Detection и Connected Components представлены на рисунках 5 и 6 соответственно.



*Рисунок 5. Community Detection*



*Рисунок 6 . Connected Components*

Как видно из рисунков, точность у DANMF гораздо выше. DANMF дал гораздо более правдивую картину о сообществах. Минусом является то, что необходимо заранее знать о количестве сообществ для получения хорошего результата, в то время как у классических методов такого недостатка нет.

В первую очередь это связано с более комплексным подходом алгоритма и наличием глубокого обучения в структуре работы. В то время как Connected Components и Community Detection заключают в себе более простую и просто реализуемую логику работы, с чем и связана их большая погрешность и некоторая неточность в результатах.

### **Список литературы:**

- 1 Easley, D. Networks, crowds, and markets: Reasoning about a highly connected world [Text] / D. Easley, J. Kleinberg. -Cambridge: Cambridge University Press, 2010. - 833 p.
- 2 Girvan, M. Community structure in social and biological networks [Text] / M. Girvan, M. Newman // Proceedings of the National Academy of Sciences. - 2002. - 9 p.
- 3 Ian, X. Y. Towards real-time community detection in large networks [Text] / X.Y. Ian, P. Pan, P. Lio, J. Crowcroft // Cambridge: Cambridge University Press. - 2008. - 11 p.
- 4 Geoffrey, E. A fast learning algorithm for deep belief nets [Text] / E. Geoffrey, S. Osindero, Y.-W. Teh // University of Toronto: Department of Computer Science -2006. -28 p.
- 5 Bengio, Y. Representation learning: A review and new perspectives [Text] / Y. Bengio, A. Courville, P. Vincent // Canadian Institute for Advanced Research: U. Montreal. - 2013. - 30 p.
- 6 Trigeorgis, G. A deep semi-nmf model for learning hidden representations [Text] / G. Trigeorgis, K. Bousmalis, B. Schuller // Proceedings of the 31st International Conference on International Conference on Machine Learning. - 2014. - 9 p.