

Постовалов Степан Сергеевич – ООО "Сигма-Групп"; e-mail: poroh20100@yandex.ru; 357500, г. Пятигорск, ул. Адмиральского, д. 2, к. 2, кв. 813; тел.: 89064996162; ведущий разработчик; магистр.

Макаров Анатолий Михайлович – Пятигорский государственный лингвистический университет; e-mail: mellin_22@mail.ru; 357532, г. Пятигорск, пр. Калинина, 9; тел.: 89283746783; кафедра информационно-коммуникационных технологий, математики и информационной безопасности; профессор.

Postovalov Stepan Sergeevich – "Sigma Group" LLC; e-mail: poroh20100@yandex.ru; 357500, Pyatigorsk, Admiral'skogo ыекуе, 2, k. 2, room 813; phone: +79064996162; lead developer; master's.

Makarov Anatoliy Mikhaylovich – Pyatigorsk State Linguistic University; e-mail: mellin_22@mail.ru; 357532, Pyatigorsk, Kalinina Ave., 9; phone: +79283746783; the department of information and communication technologies, mathematics and information security; professor.

УДК 004.891.2

DOI 10.23683/2311-3103-2018-3-33-45

А.Н. Целых, В.С. Васильев, Л.А. Целых

КЛАСТЕРИЗАЦИЯ ОРИЕНТИРОВАННЫХ ВЗВЕШЕННЫХ ЗНАКОВЫХ ГРАФОВ НА ОСНОВЕ ФУНКЦИОНАЛА ПОТЕНЦИАЛЬНОЙ ЭНЕРГИИ УПРУГОЙ ДЕФОРМАЦИИ С ИСПОЛЬЗОВАНИЕМ КОГНИТИВНЫХ МОДЕЛЕЙ*

Предлагается новый подход к кластеризации на основе модели разделения вершин на группы по признаку более высокой потенциальной энергии ребер внутри групп, чем между ними. Этот метод реализуется посредством минимизации функционала потенциальной энергии упругой деформации для ориентированных взвешенных знаковых графов. Существующие структуры в матрице смежности графа выражаются в ее структурированности. Поиск наилучшего упорядочения вершин графа производится посредством процесса оптимизации в пространстве действительных чисел, что неизбежно отображается на множество перестановок индексов вершин. Метод разрабатывался для сетей, представляющих взаимоотношения в экономических системах. Эти взаимоотношения представлены разнородными факторами и причинно-следственными связями между ними. Система объектов и их взаимоотношения представлены в виде нечеткой когнитивной карты, которая, по сути, является ориентированным взвешенным знаковым графом. К этому графу и применяются методы кластеризации. Новизна подхода заключается в том, что решение задачи кластеризации находится как решение задачи оптимизации функции многих переменных (функционалов). Предложенный метод использует механистическую аналогию, которая является частным случаем метрических представлений, что обеспечивает выпуклость задачи. Данный подход позволяет конструировать различные функционалы с понятной интерпретацией и предсказуемостью процесса минимизации. Работа алгоритма заключается в нахождении нумерации вершин графа, при которой достигается наименьшее значение функционала. На этой нумерации компоненты градиента функционала используются для определения границ кластеров. Критерием отнесения вершин к одному сообществу является близость индексов и одинаковые знаки градиента. Эти критерии являются четко формализуемыми. Предлагаемый алгоритм относит вершины к тому или иному кластеру и, одновременно, определяет порядок вершин в кластере. Этот порядок отражает степень распределение вершин по соотношению внутрикластерной энергии и внекластерной энергии связей. Иерархическая структура графа выявляется путем рекурсивного применения предлагаемого алгоритма внутри каждого кластера, игнорируя межкластерные связи. Применяемый в алгоритме кластеризации функ-

* Данная работа произведена при поддержке гранта РФФИ № 16-01-00098.

ционал потенциальной энергии упругой деформации отражает причинно-следственный характер взаимоотношений между факторами в социально-экономической системе. Минимизация функционала происходит монотонно и не требует вмешательства пользователя. Алгоритм является вычислительно эффективным.

Кластеризация; функционал упругой деформации; ориентированный взвешенный граф; оптимизационные методы; когнитивные карты.

A.N. Tselykh, V.S. Vasilev, L.A. Tselykh

CLUSTERING OF ORIENTED WEIGHTED GRAPHS BASED ON THE FUNCTIONAL OF POTENTIAL ENERGY OF ELASTIC DEFORMATION WITH THE USE OF COGNITIVE MODELS

A new approach to clustering based on the model of division of vertices into groups based on higher potential energy of edges within groups than between the groups themselves is proposed. This method is implemented by minimizing the elastic strain potential energy functional for oriented weighted signed graphs. Existing structures in the adjacency matrix of a graph are expressed in its structure. The search for the best ordering of the graph vertices is carried out by means of the optimization process in the space of real numbers, which is inevitably displayed on the set of permutations of vertex indices. The method is developed for networks representing inter-relations in economic systems. These relations are represented by heterogeneous factors and cause-effect relationships between them. Systems of objects and their relations are presented in the form of a fuzzy cognitive map, which, in fact, is a weighted oriented sign graph. Clustering methods are applied to this graph. The novelty of the approach is that the solution of the clustering problem is found as a solution to the optimization problem of the function of many variables (functionals). The proposed method uses a mechanical analogy, which is a special case of metric representations, which provides the convexity of the problem. This approach allows us to design various functionalities with a clear interpretation and predictability of the minimization process. The work of the algorithm is to find the numbering of the vertices of the graph, which reaches the lowest value of the functional. In this numbering, gradient components of the functional are used to determine the boundaries of the clusters. The criterion for classifying vertices as a single community is the proximity of the indices and the same gradient signs. These criteria are clearly formalized. The proposed algorithm classifies the vertices of a particular cluster and at the same time determines the order of the nodes in the cluster. This order reflects the degree of vertex distribution in relation to the intracluster energy to the energy of the extracuster bond. The hierarchical structure of the graph is revealed by recursive application of the proposed algorithm in each cluster without taking into account inter-cluster connections. The potential energy of the elastic deformation functional used in the clustering algorithm reflects the causal nature of the relationship of factors in the socio-economic system. Functional minimization is monotonous and requires no user intervention. The algorithm is computationally efficient.

Clustering; elastic deformation functional; oriented weighted graph; optimization methods; cognitive maps.

Введение. Разработке методов выявления сообществ в сетях реального мира, таких как социальные и природные сети, посвящено много исследований. Исследования по выявлению кластеров в экономических системах, в основном, касались задач распространения информации (например, маркетинговой), решения технических задач в экономических системах (например, в производственных системах для распределения нагрузок потребителей электроэнергии [1], классификации клиентов и потребителей [2], задач обслуживания [3] группировки и классификации экономических и финансовых данных [4], также и в сочетании экспертных оценок состояний переменных с эмпирическими данными [5] и др.

Общим для всех перечисленных подходов является использование, как правило, набора данных и метрических переменных. Используемые наборы данных представлены небольшим количеством переменных и не являются системным представлением исследуемой области, а только отражают отдельные характери-

стики системы относительно сфер влияния. В данном исследовании будут рассматриваться управленческие задачи, имеющие преимущественно качественные характеристики объекта исследования. Представление предметной области в виде целостной системы и использование качественных характеристик сделают результаты более объективными. Основным отличием сетей, представляющих взаимоотношения в экономических системах, от других сетей является отсутствие ясно видимых связей между акторами системы. Например, в социальных сетях в Интернет как акторы, так и связи присутствуют в явном виде. Акторами всегда являются пользователи сети. Связи присутствуют в виде различных links, sms, поддерживающих лайков, разных статистик заходов на сайт и тому подобных явно выражаемых и фиксируемых взаимоотношений между акторами. В природных, биологических и технических системах также имеются акторы и связи между ними в виде, возможном для определения и фиксации. Напротив, в экономических системах как акторы, так и связи не носят явного характера. Поэтому требуется инструмент, позволяющий сначала представить саму систему объектов и их взаимоотношений в виде графа, к которому потом можно применить методы кластеризации. Таким инструментом, на наш взгляд, является техника когнитивных карт, предложенная Kosko в 1986 году [6]. Нечеткие когнитивные карты представляют собой ориентированные взвешенные знаковые графы. Поэтому к ним могут быть применены соответствующие методы кластеризации [7]. Однако специфика экономических взаимоотношений накладывает ограничения на их выбор и применение. В общем случае отношения в социальных сетях между концептами системы носят поведенческий характер. В экономических системах отношения носят принудительный характер, изменение одного фактора обязательно влечет за собой изменение другого. Кроме этого сами факторы экономической системы являются неоднородными, имеют различные характеристики, и, как правило, не схожие параметры. Когнитивные карты выражают причинно-следственные отношения между факторами системы, что соответствует реальным взаимоотношениям.

Тогда решаемая проблема в настоящем исследовании формулируется следующим образом – нахождение кластеров в ориентированном взвешенном графе с учетом причинно-следственных связей, сила влияния которых выражена весами на ребрах. В этом исследовании предлагается алгоритм, который обнаруживает иерархические непересекающиеся структуры в ориентированном взвешенном графе.

1. Связанные работы. Рассмотрим некоторые методы кластеризации, чтобы подчеркнуть возможность их применения для решения нашей проблемы и определить их различия с предложенным подходом. Также акцентировано внимание на проблемах, связанных с формализацией задачи, таких как, направленность ребер графа и содержательная значимость веса на ребрах графа. Было проанализировано большое количество статей, рассматривающих методы кластеризации. Приведенные в этом разделе исследования, во-первых, использовали понятие качества связи между вершинами при разработке алгоритма кластеризации, например, влияние; во-вторых, приведенный в них порядок вершин получил наиболее близкие значения функционала упругой энергии по сравнению с нашими значениями функционала.

Обширное исследование методов кластеризации в направленных графах представлено в survey авторов Malliarosa & Vazirgiannis [7]. Приведено описание проблем, возникающих в связи с несимметричностью матриц смежности, представляющих граф, и подходов к их симметризации.

Огромная исследовательская работа по обзору методов обнаружения сообществ в графах представлена S. Fortunato [10]. В этой работе также рассмотрены предложенные Nepusz et al. [11] итеративные градиентные методы оптимизации на основе минимизации функции стоимости, использующие расстояния между точ-

ками в пространстве. Относительно проблемы, поставленной в нашем исследовании, данный подход имеет возможности обобщения на различные интерпретации смыслового содержания функции стоимости для различных задач. В другой работе S. Fortunato [12] представлен критический анализ, посвященный вопросам валидации алгоритмов кластеризации и их сравнительным характеристикам.

В работе [13] авторы предлагают двухэтапный алгоритм CNNI с функцией оценивания на основе коэффициента влияния λ . Условия значения λ не определены и предполагают наличие начальных представлений о принадлежности узла к кластеру. В задачах с неизвестными представлениями о кластерах метод содержит невыполнимые условия. Алгоритм рассчитан на невзвешенные графы.

Алгоритм, основанный на понятии ближайших соседей и принципе суперпозиции для влияния, предлагается в статье [14]. Используемая в алгоритме дистанция не применима для ориентированных графов. Кроме того, метод рассчитан на использование специфических данных, характеризующихся однотипными координатами. Эти специфические особенности применения алгоритма существенно ограничивают его применение.

Алгоритм, разработанный в исследовании Šubelj & Bajec [15], основан на технике распространения меток вершин и рассчитан на взвешенные, но ненаправленные графы. С помощью единственного весового параметра производится управление приоритетом в определении групп вершин или модулей вершин. Однако сочетание техник Label propagation и random walks накладывает ограничения на присваивание ребрам весов. В направленном графе процесс распространения меток может не обладать сходимостью. Например, в случае наличия циклов.

Разработанная Min, Yu, & Li [16] новая техника взвешивания предполагает усреднение весов дуг на маршруте длины k , что возможно в случаях, когда веса всех дуг имеют единый смысл. Например, для отношений в соцсетях. Для весов, отражающих причинно-следственные связи, такое усреднение недопустимо.

В работе [17] авторы предлагают алгоритм на основе магистральной степени и экспансии для невзвешенных ненаправленных сетей. Применение предложенных определений к взвешенным сетям может существенным образом изменить порядок ребер в результате сортировки, что повлияет на определение сообществ. При этом изменяется также смысловое содержание модели, положенное в основу алгоритма.

Алгоритм, предложенный в [18], применим к графам с симметричной матрицей смежности. Для каждой вершины определяются функция Грина для уравнения Пуассона, описывающего возникающие перепады давления на ребрах графа. Попадание в одно сообщество определяется по степени схожести функций Грина для этих вершин. Принадлежность вершин к сообществу обладает транзитивностью, а меры схожести функции Грина, оцениваемые Евклидовой нормой, транзитивностью не обладают. Так же, как не обладает транзитивностью понятие корреляции. В целом, данный подход в наибольшей степени отражает условия нашей проблемы.

2. Методы. Гипотеза, рассматриваемая для разработки алгоритма кластеризации: так как эффективные методы оптимизации функций многих переменных изменяют значения сразу всех координат, то ожидается эффект от применения таких методов к задачам оптимизации на графах, где обновленная нумерация охватит сразу все вершины.

2.1. Математическая постановка задачи кластеризации. Рассмотрим конечный граф $G = \langle V, E \rangle$, где $V = \{v_1, v_2, \dots, v_n\}$ есть конечное множество вершин, n есть количество вершин, $E = \{\langle v_i, v_j \rangle | v_k \in V, v_1 \in V\}$ есть конечное множество дуг. Графу G соответствует матрица смежности $A = \|a_{ij}\|_{n \times n}$, где вес a_{ij} дуги $\langle v_i, v_j \rangle$ может выражать наличие дуги (булево значение), кратность дуги, вес дуги,

нечёткую меру смежности вершин v_i и v_j и др. Применительно к решению оптимизационных задач на графе G в пространстве R_n имеют значение точки (x_1, x_2, \dots, x_n) с целочисленными координатами, соответствующие перестановкам $\langle i_1, i_2, \dots, i_n \rangle$ множества индексов вершин $\langle 1, 2, \dots, n \rangle$.

Существующие в графе структуры выражаются в структурированности матрицы смежности. Портрет матрицы смежности зависит от последовательности нумерации вершин. При изменении порядка $\langle 1, 2, \dots, n \rangle$ нумерации вершин на $\langle i_1, i_2, \dots, i_n \rangle$ будет изменяться и портрет матрицы смежности. Выявление структур в портрете матрицы смежности означает выявление структур в исходном графе. Если выполнить непрерывное продолжение множества перестановок $\langle i_1, i_2, \dots, i_n \rangle$ индексов вершин на множество действительных чисел $(x_1, x_2, \dots, x_n) \in R_n$, то для поиска наилучших упорядочений можно применить методы оптимизации функций нескольких действительных переменных $\Phi(x_1, x_2, \dots, x_n)$ [21]. При этом в общем случае, в целевой функции оптимизационной задачи могут быть учтены не только исходные данные о самом графе, но и информация о достигнутом на очередной итерации взаимном расположении вершин. Следует иметь в виду, что изначально речь идёт о поиске наилучшего упорядочения вершин графа, поэтому процесс оптимизации в пространстве действительных чисел должен неизбежно отображаться на множество перестановок индексов вершин. Необходимое отображение действительно-значных решений (x_1, x_2, \dots, x_n) на множество перестановок $\langle i_1, i_2, \dots, i_n \rangle$ будет проявлять себя «шоками», способными изменить соотношение производительностей алгоритмов. Тем не менее, ставится задача удерживать получаемые алгоритмы в границах $O(n^3)$ затрат.

Обобщением изложенного подхода будет его применение не только к задачам квадратичного программирования, но и к задачам минимизации функционалов достаточного общего вида [19]. Существование решения задачи минимизации будет обеспечиваться не ограниченностью снизу значений функционалов, а конечной мощностью множества перестановок индексов вершин графа G . Другими словами, фактически будут решаться задачи условной оптимизации. Желательным требованием будет выпуклость минимизируемых функционалов, хотя бы нестрогая, для единственности решений рассматриваемых задач и для независимости результатов процессов оптимизации от выбора начальных приближений.

Новизна подхода заключается в том, что решение задачи кластеризации находится как решение задачи оптимизации функции многих переменных (функционалов), где в отличие от чисто топологических методов кластеризации предложен метод, использующий метрические представления, обеспечивающие выпуклость задачи. Среди метрических представлений механистическая аналогия, являющаяся частным случаем, позволяет конструировать различные функционалы с понятной интерпретацией, предсказуемостью процесса минимизации, допускающим гибкое управление и реализацию эффективными методами.

В пространстве R_n действительных переменных x_j ($1 \leq j \leq n$) рассматриваются функционалы $\Phi(x_1, x_2, \dots, x_n)$. Функционалы могут, но не обязаны быть квадратичными. Однако даже для нелинейных неквадратичных функционалов известные эффективные численные методы оптимизации оперируют матрицей вторых производных, которая в случае квадратичных функционалов является неизменной, а в случае нелинейных неквадратичных – зависит от точки $(x_1, x_2, \dots, x_n) \in R_n$. Квадратичные функционалы можно представить в виде

$$\Phi(x_1, x_2, \dots, x_n) = \sum_{i=1}^n \sum_{j=1}^n x_i \left(\frac{1}{2} s_{ij} x_j - f_i \right) = \Phi(\mathbf{x}) = \frac{1}{2} (\mathbf{S}\mathbf{x}, \mathbf{x}) - (\mathbf{f}, \mathbf{x}), \quad (1)$$

где $\mathbf{S} = \|s_{ij}\|_{n \times n}$ – матрица вторых производных; $\mathbf{f} = (f_1, f_2, \dots, f_n)^T$ – вектор коэффициентов линейной компоненты; $\mathbf{x} = (x_1, x_2, \dots, x_n)^T$ – вектор координат точки.

Среди квадратичных функционалов основным рассматриваемым будет

$$\Phi(x_1, x_2, \dots, x_n) = \frac{1}{2} \sum_{i=1}^n \sum_{j=1}^n a_{ij} (x_j - x_i)^2. \quad (2)$$

Функционал (1) можно интерпретировать как потенциальную энергию упругой деформации механической системы [20], где a_{jk} является жёсткостью пружины, а $|x_j - x_k|$ – её растяжением. В силу одного из основных принципов физики, механическая система стремится к минимуму своей потенциальной энергии. В результате минимизации функционала (1) вершины, связанные дугами с большей «пружинной жёсткостью», должны получить в порядке индексов $\langle i_1, i_2, \dots, i_n \rangle$, по возможности, соседние индексы. Тем самым будут выявлены или кластеры, или причинно-следственные связи (или одновременно и то, и другое, в зависимости от решаемой задачи). Представим функционал (2) в виде:

$$\Phi(x_1, x_2, \dots, x_n) = \frac{1}{2} \sum_{i=1}^n \sum_{j=1}^n a_{ij} (x_j - x_i)^2 = \sum_{i=1}^n x_i^2 \sum_{j=1}^n \frac{1}{2} (a_{ij} + a_{ji}) - \sum_{i=1}^n \sum_{j=1}^n a_{ij} x_i x_j.$$

Соотношение между элементами матрицы смежности и элементами матрицы квадратичной формы следующее:

$$s_{ij} = \begin{cases} -a_{ii} + \sum_{j=1}^n \frac{1}{2} (a_{jk} + a_{kj}), & l = k. \\ -a_{kl}, & l \neq k \end{cases}$$

2.2. Алгоритм кластеризации на основе функционала упругой энергии. Задача безусловной минимизации квадратичного функционала (1) эквивалентна задаче решения системы линейных алгебраических уравнений $\frac{1}{2}(\mathbf{S} + \mathbf{S}^T)\mathbf{x} = \mathbf{f}$.

Для ее решения используем метод скорейшего спуска [19]. Зададим начальное приближение $(x_1^{(0)}, x_2^{(0)}, \dots, x_n^{(0)}) = (1, 2, \dots, n)$, где верхний индекс $^{(0)}$ у переменных x означает номер шага итерационного процесса. Следующее приближение после шага $q = 0; 1; \dots$ вектора координат точки $\mathbf{x}^{(q+1)} = (x_1^{(q+1)}, x_2^{(q+1)}, \dots, x_n^{(q+1)})^T$ в методе скорейшего спуска определяется как доставляющее минимум функционалу (1) вдоль направления $\mathbf{d}^{(q)} = (d_1^{(q)}, d_2^{(q)}, \dots, d_n^{(q)})^T$, что выражается следующим образом:

$$\mathbf{x}^{(q+1)} = \mathbf{x}^{(q)} + t_q \mathbf{d}^{(q)},$$

где $x^{(q)}$ есть приближение к решению на шаге q , t есть возможная длина шага, $t_q = \arg \min \Phi(\mathbf{x}^{(q)} + t\mathbf{d}^{(q)})$ есть оптимальное значение шага.

Получим

$$\Phi(\mathbf{x}^{(q+1)}) = \frac{1}{2}(\mathbf{S}\mathbf{x}^{(q+1)}, \mathbf{x}^{(q+1)}) - (\mathbf{f}, \mathbf{x}^{(q+1)}) = \min \left(\frac{1}{2}(\mathbf{S}(\mathbf{x}^{(q)} + t\mathbf{d}^{(q)}), \mathbf{x}^{(q)} + t\mathbf{d}^{(q)}) - \mathbf{f}, \mathbf{x}^{(q)} + t\mathbf{d}^{(q)} \right) = \frac{1}{2}(\mathbf{S}\mathbf{x}^{(q)}, \mathbf{x}^{(q)}) - (\mathbf{f}, \mathbf{x}^{(q)}) + t(\mathbf{S}\mathbf{d}^{(q)}, \mathbf{x}^{(q)}) - (\mathbf{f}, \mathbf{d}^{(q)}) + \frac{t^2}{2}(\mathbf{S}\mathbf{d}^{(q)}, \mathbf{d}^{(q)}) - t(\mathbf{f}, \mathbf{d}^{(q)}).$$

откуда в общем случае (1) с учётом того, что в методе скорейшего спуска

$$\mathbf{d}^{(q)} = -\nabla\Phi(\mathbf{x}^{(q)}),$$

$$t_q = \left(\mathbf{f} - \frac{1}{2}(\mathbf{S} + \mathbf{S}^T)\mathbf{x}^{(q)}, \mathbf{d}^{(q)} \right) / (\mathbf{S}\mathbf{d}^{(q)}, \mathbf{d}^{(q)}) = -\nabla\Phi(\mathbf{x}^{(q)}, \mathbf{d}^{(q)}) / (\mathbf{S}\mathbf{d}^{(q)}, \mathbf{d}^{(q)}) = (\mathbf{d}^{(q)}, \mathbf{d}^{(q)}) / (\mathbf{S}\mathbf{d}^{(q)}, \mathbf{d}^{(q)}),$$

а применительно к функционалу (2) компоненты градиента функционала выражаются:

$$\Phi^{(q)} = \frac{1}{2} \sum_{k=1}^n \sum_{l=1}^n a_{kl} (x_l^{(q)} - x_k^{(q)})^2,$$

$$-\nabla \Phi(x^{(q)}) = \frac{\partial \Phi^{(q)}}{\partial x_j} = -d_j^{(q)} = -\sum_{k=1}^n (a_{jk} + a_{kj})(x_k^{(q)} - x_j^{(q)}), j = 1, \dots, n,$$

n – есть индекс вершины.

$$t_q = -\sum_{k=1}^n \sum_{l=1}^n a_{kl} (x_l^{(s)} - x_k^{(s)})(d_l^{(s)} - d_k^{(s)}) / \sum_{k=1}^n \sum_{l=1}^n a_{kl} (d_l^{(s)} - d_k^{(s)})^2,$$

При вычислении компонент градиента происходит симметризация symmetrization типа $\mathbf{A} + \mathbf{A}^T$, согласно классификации, предложенной в работе [7]. Задача инварианта относительно циклической перестановки вершин, поэтому условием окончания процесса минимизации методом скорейшего спуска является выход в одну из равнозначных вершин минимума функционала.

Для решения задачи минимизации квадратичных функционалов и отображения их в множество перестановок целочисленных индексов вершин графа сортировкой новых значений переменных предлагается следующий алгоритм.

Алгоритм. Минимизация квадратичного функционала Φ градиентным методом скорейшего спуска:

1. Ввод элементов матрицы смежности $a_{i,j}$, $1 \leq i, j \leq n$, порога ε и задание начальной перестановки $\mathbf{v}^{(0)}$ (напр., $\mathbf{v}^{(0)} = \langle 1, 2, \dots, n \rangle$).

2. Выполнение функции $mf(i, m)$.

2.1. Вычисление компонент $\nabla_i \Phi^{(s)}(\mathbf{v}^{(s)})$, $1 \leq i \leq n$ градиента функционала $\Phi^{(q)}$.

2.2. Определение шага t_s .

2.3. Определение новых значений вещественных переменных $x_i^{(s+1)} = x_i^{(s)} - t_s \nabla_i \Phi^{(s)}$, $1 \leq i \leq n$.

2.4. Сортировка вещественных переменных $x_i^{(s+1)}$, $1 \leq i \leq n$ и получение целочисленной перестановки $\mathbf{v}^{(s+1)} = \langle i_1^{(s+1)}, i_2^{(s+1)}, \dots, i_n^{(s+1)} \rangle$.

2.5. Повторение 2.1.–2.4. пока не будет достигнуто заикливание $\mathbf{v}^{s+p} = \langle v_1^{s+p}, v_2^{s+p}, \dots, v_n^{s+p} \rangle = \langle v^s \rangle = \langle v_1^s, v_2^s, \dots, v_n^s \rangle$, где $s > T, p > 0$. Выполнение для всех шагов s , следующих после шага T , где p есть период повторяемости.

3. Вывод перестановки $\mathbf{v}^* = \langle i_1^*, i_2^*, \dots, i_n^* \rangle$, доставляющей минимум функционалу Φ .

2.3. Определения.

Определение 1. Минимум потенциальной энергии упругой деформации:

$$\Phi_{min} = \frac{1}{2} \sum_{i=1}^n \sum_{j=1}^n a_{ij} (v_j - v_i)^2,$$

где i, j есть индексы вершин в порядке их нумерации; a_{ij} есть вес дуги, направленной от вершины i к вершине j ; v_i, v_j есть вершины графа соответственно с индексами i и j в порядке их нумерации; n есть количество вершин графа.

В качестве метрики оценки степени кластеризации принято значение самого функционала (1). Это значение отражает положение относительного предельного равновесия системы (статическую конфигурацию), достигнутое при минимуме функционала потенциальной энергии упругой деформации.

Определение 2. Индекс потенциальной энергии вершины в общей системе есть компонента градиента функционала:

$$\varphi_i = \frac{1}{2} \sum_{j=1}^n (a_{ij} + a_{ji})(v_i - v_j)^2, 1 \leq i \leq n.$$

где i, j — это индексы вершин в порядке их нумерации; a_{ij} — это вес дуги, направленной от вершины i к вершине j ; a_{ji} — это вес дуги, направленной от вершины j к вершине i ; v_i, v_j — это вершины графа соответственно с индексами i и j в порядке их нумерации; n — это количество вершин графа.

Определение 3. Градиент потенциальной энергии упругой деформации указывает направление наиболее быстрого роста функционала, тогда i -ая компонента градиента выражается в виде:

$$\nabla_i \Phi = \sum_{j=1}^n (a_{ij} + a_{ji})(v_i - v_j), \quad 1 \leq i \leq n.$$

где i, j — это индексы вершин в порядке их нумерации; a_{ij} — это вес дуги, направленной от вершины i к вершине j ; a_{ji} — это вес дуги, направленной от вершины j к вершине i ; v_i, v_j — это вершины графа соответственно с индексами i и j в порядке их нумерации; n — это количество вершин графа.

Определение 4. Порядок вершин в кластере выражается в виде

$$\langle v[i]; v[i + 1]; \dots; v[m] \rangle.$$

Порядок отражает степень распределение вершин по соотношению внутрикластерной и внекластерной энергии связи.

Определение 5. Две соседние вершины относятся к разным кластерам, если происходит смена знака у соответствующих им компонент градиента функционала со знака «плюс» на знак «минус», т.е. с $\nabla \Phi_i > 0$ на $\nabla \Phi_{i+1} < 0$, где i — это номер вершины, входящей в текущий кластер, $i + 1$ — это номер вершины, входящей в следующий кластер.

Это означает, что вершина попадает в кластер, если уменьшение потенциальной энергии от приближения к другим вершинам кластера преобладает над увеличением потенциальной энергии из-за удаления от вершин, не входящих в кластер.

3. Эксперимент. Численный эксперимент проведен для известных эталонных задач кластеризации социальных сетей Zachary Karate Club, Bottlenose Dolphins, American College football networks, а также задач реальных данных разной размерности и сложности.

3.1. Численный эксперимент на эталонных графах.

а) *Zachary Karate Club network.* В соответствии с реальными данными из [22], 34 члена клуба каратэ разделились на 2 группы. Одна группа (17 человек) осталась с администратором клуба Mr. John A., другая (17 человек) покинула клуб вместе с инструктором клуба Mr. Hi. Предлагаемый метод на основе минимизации функционала разделил граф на два кластера. Расчеты производились для двух матриц: невзвешенной и взвешенной. Как и ожидалось, применение взвешенной матрицы смежности, рассчитанной на основе взаимодействия членов клуба между собой, показали высокую степень точности разбиения с $NMI=1$. Кластеризация невзвешенной матрицы дала менее точные результаты, меняя местами вершины 8 и 40. Процесс минимизации в процессе реализации предлагаемым алгоритмом представлен на рис. 1.

б) *Bottlenose Dolphins network.* Следующая социальная сеть, рассмотренная для тестового примера, является стаей дельфинов-афалин, обитающих в Doubtful Sound, New Zealand [23, 24] в состав которой входят 64 дельфина. Стая разделилась на 2 группы, когда один ключевой дельфин по имени SN100 временно выбыл. Отделившаяся группа дельфинов вскоре возвратилась с возвратом дельфина SN100.

Представляемый здесь алгоритм идентифицирует два сообщества так же, как и Girvan-Newman алгоритм [25]. Процесс минимизации в процессе реализации предлагаемым алгоритмом представлен на рис. 2.

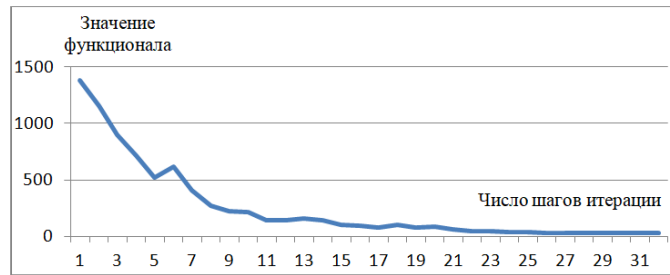


Рис. 1. Значения функционала в процессе кластеризации сети Zachary Karate Club

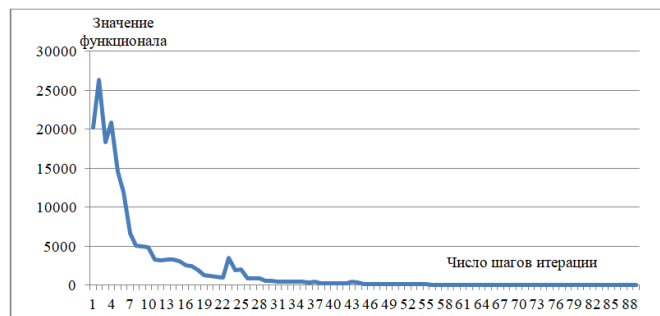


Рис. 2. Значения компонент функционала в процессе кластеризации сети дельфинов-афалин

Процесс кластеризации сети дельфинов-афалин был осуществлен в один этап, на котором произошло разделение графа на 5 кластеров (рис. 3). Значения показателя NMI рассчитывались в сравнении с результатами, полученными в работе [24], и с результатами, полученными в статье [25], и составили 0,8559 и 0,86 соответственно.

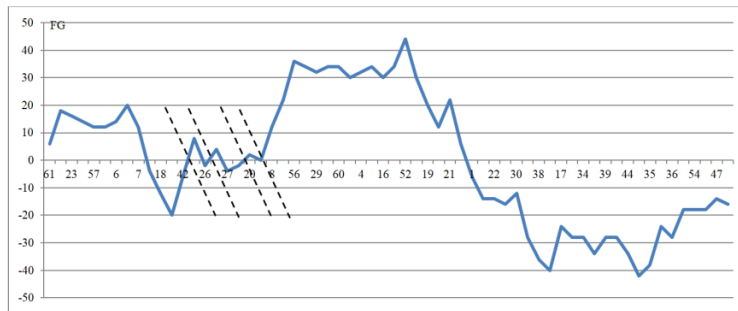


Рис. 3. Значения компонент градиента функционала по результатам кластеризации сети дельфинов-афалин; пунктирные линии показывают раздел на кластеры

3.2. Численный эксперимент на графах, представляющих когнитивные модели. Далее рассмотрены результаты кластеризации известной когнитивной модели устойчивости банкинга авторов Ferreira et al. [26]. Минимизация функционала производилась в два этапа. При этом значение функционала минимизировалось от 88046,9 до 23,2 (рис. 4).

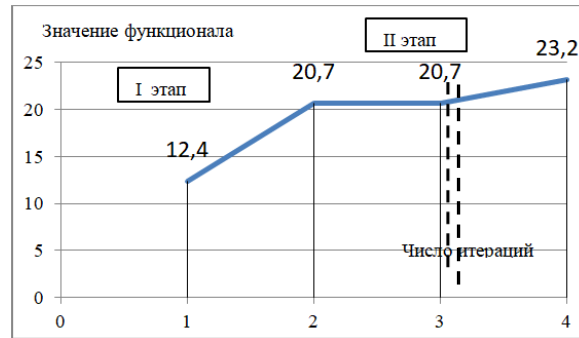


Рис. 4. Значения функционала в процессе кластеризации графа модели устойчивости банкинга

На первом этапе были получены 5 кластеров, три из которых на втором этапе окончательно разделились на субкластеры. В целом по результатам кластеризации было получено 8 кластеров, полностью соответствующих реальному разбиению ($NMI=1$). Как показано на рис. 5 в графе имеются вершины с высоким значением компоненты градиента функционала. Эти вершины образуют явно выраженные центры кластеров с наивысшим значением, отмечаемые и в работе [26]. График потенциальной упругой энергии системы до минимизации функционала и после (рис. 5) показывает явное успокоение системы, выраженное в упорядочении и сглаживании графика, что находится в консенсусе со смыслом этого функционала. То есть система стремится принять некоторое предельное равновесное положение (статическую конфигурацию) с возможным минимумом суммарной потенциальной энергии в целом и внутри кластеров. Кроме этого, среди внутренних структур графа могут быть выделены различные характерные конфигурации. Индексы вершин φ_i являются очень хорошим индикатором выявления центров звездных подграфов. На рис. 5 видны четко выраженные пики, соответствующие таким центрам.

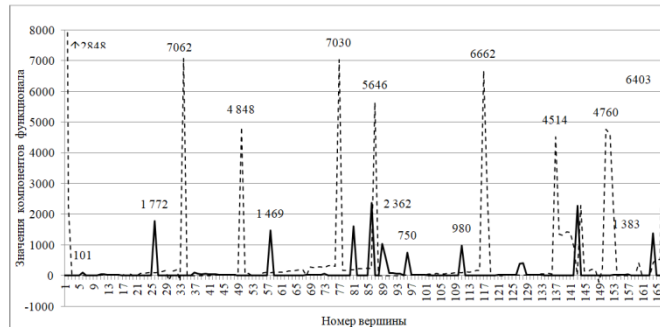


Рис. 5. Потенциальная энергия упругой деформации системы: пунктирная линия – до кластеризации; сплошная линия – после кластеризации

3.3. Оценка результатов экспериментов. Результаты кластеризации предлагаемым методом сравнивались по значению функционала, достигнутого для порядка вершин на дендрограммах, полученных при кластеризации другими известными методами. Для сравнения были взяты дендрограммы из следующих статей для тестовых сетей: [13, 15–18, 25, 27–29, 30–32]. Анализ результатов показывает, что наиболее близкие значения функционалов получены по дендрограммам из работ, в которых методы кластеризации предназначены для взвешенных сетей и

имеют в своей основе физическую аналогию, как, например, в [18]. Время вычислений предлагаемым алгоритмом на исследованных графах находится в пределах $O(n)$ затрат и не превышает 50 sec.

4. Обсуждение результатов. Оценка алгоритма произведена на основе следующих критериев и выявила следующее:

а) Применимость для предметной области. Проведенные исследования показали применимость данного подхода с использованием функционалов на основе упругой энергии для анализа когнитивных моделей, представляющих собой ориентированные взвешенные знаковые графы, отображающие причинно-следственные связи на ребрах.

б) Реализуемость. Функционал является знакоопределенным, т.е. ограничен снизу нулевыми значениями, выпуклыми в силу доказательств теоремы. Поэтому процесс минимизации, в том числе и в постановке задачи безусловной оптимизации, имеет решение, достижимое любыми методами 0-го и 1-го порядка. В отличие от целочисленных алгоритмов, сложность оценивается как NP. В предлагаемом алгоритме на каждой итерации каждая переменная получает новое значение. Это существенно ускоряет его сходимость, и всегда закончится в оптимальной точке.

в) Формат кластеров. Предложенный алгоритм не требует заранее определенного количества кластеров, получая кластеры «как есть».

г) Качество полученных кластеров. Качество получаемых кластеров оценивается на точность и верность отнесения вершин в кластеры, а также наличие порядка вершин в них.

Оценка качества полученных кластеров произведена на тестовых моделях реального мира (Karate Club, Dolphins, Football и ряда других) и более 100 когнитивных моделях. В статье приведены результаты кластеризации тестового графа, представляющего известную когнитивную модель устойчивости банковского бизнеса авторов Ferreira et.al. [26]. NMI составляет 1.

Результаты кластеризации также могут быть оценены по точности и верности с точки зрения их состава (конкретного состава вершин графа) экспертами предметной области. Результаты кластеризации графа тестовой когнитивной модели показали полное соответствие их содержательному контенту.

Кроме этого, предлагаемый алгоритм не только относит вершины к тому или иному кластеру, но также определяет порядок вершин в кластере. Этот порядок отражает степень распределение вершин по соотношению внутрикластерной и внекластерной энергии связи.

д). Производительность. Время решения задачи для матриц размером $10^3 \times 10^3$ находится в пределах 20-50 sec. Минимизация функционала происходит монотонно и не требует вмешательства пользователя. Алгоритм является вычислительно эффективным. На исследованных примерах алгоритм реализовал себя не хуже, чем $O(n)$ затрат.

Заключение. В этом исследовании мы фокусируемся на проблеме обнаружения иерархических структур непересекающихся сообществ в ориентированных взвешенных знаковых графах, представляющих социально-экономическую систему. Такое представление системы является ключевым инструментом для понимания взаимоотношений факторов системы и ее структуры.

Наш алгоритм относится к классу методов разделения (partitional methods). Иерархия генерируется с помощью дивизионального подхода и основана на группе субкластеров, созданных методом разделения. Производительность нашего метода влияет размерность графа. Время вычисления пропорционально размеру сети и является критичным, начиная с размерности 10^3 . Реализация этого метода не чувствительна к распределению набора данных, пользовательская настройка и

регулировка не требуется. Эксперименты показали, что алгоритм удовлетворительно эффективно обнаруживает структуры в тестовых социальных сетях, а также в тестовых графах, представляющих когнитивные карты. Эти сильные стороны делают этот метод широко применимым в реальных приложениях с когнитивными моделями.

БИБЛИОГРАФИЧЕСКИЙ СПИСОК

1. Biscarri F., Monedero I., García A., Guerrero J. and Leon C. Electricity clustering framework for automatic classification of customer loads, *Expert Systems With Applications*, 2017, Vol. 86, pp. 54-63.
2. Tsai C.-F., Hu Y.-H. and Lu Y.-H. Customer segmentation issues and strategies for an automobile dealership with two clustering techniques, *Expert System*, 2015, Vol. 32, No. 1, pp. 65-76.
3. Cruz A.M. Evaluating record history of medical devices using association discovery and clustering techniques, *Expert Systems with Applications*, 2013, Vol. 40 (13), pp. 5292-5305.
4. Dias J.G. and Ramos S.B. Dynamic clustering of energy markets: An extended hidden markov approach," *Expert Systems with Applications*, 2014, Vol. 41 (17), pp. 7722-7729.
5. Lorentz H., Hilmola O., Malmsten J. and Srari J.S. Cluster analysis application for understanding SME manufacturing strategies, *Expert Systems with Applications*, 2016, Vol. 66, pp. 176-188.
6. Kosko B. Fuzzy cognitive maps, *Int. Journal man-machine Studies*, 1986, Vol. 24, no. iss. 1, pp. 65-75.
7. Malliarosa F. and Vazirgiannis M. Clustering and community detection in directed networks: A survey, *Physics Reports*, 2013, Vol. 533, pp. 95-142.
8. Landau L. and Lifshitz E. Theory of Elasticity (3rd ed.). Oxford: Butterworth Heinemann, 1986.
9. Tselykh A., Vasilev V. and Tselykh L. Fuzzy graphs clustering with quality relations functionals in cognitive models, in *Advances in Intelligent Systems and Computing*, 2016.
10. Fortunato S. Community detection in graphs, *Physics Reports*, 2010, Vol. 486, p. 75-174.
11. Nepusz T., Petroczi A., Negyessy L. and Bazsy F. Fuzzy communities and the concept of bridgeness in complex networks, *Phys. Rev.*, 2008, Vol. 77, no. 1.
12. Fortunato S. and Hric D. Community detection in networks: A user guide, *Physics Reports*, 2016, Vol. 659, pp. 1-44.
13. Mu C., Liu Y., Liu Y., Wu J. and Jiao L. Two-stage algorithm using influence coefficient for detecting the hierarchical, non-overlapping and overlapping community structure, *Physica A*, Vol. 408, pp. 47-61.
14. Chen X. A new clustering algorithm based on near neighbor influence, *Expert Systems with Applications*, 2015, Vol. 42, pp. 7746-7758.
15. Šubelj L. and Bajec M. Group detection in complex networks: an algorithm and comparison of the state of the art, *Physica A*, 2014, Vol. 397, pp. 144-156.
16. Min D., Yu K. and Li H.-J. Refinement of the community detection performance by weighted relationship coupling., *Pramana - Journal of Physics*, 2017, Vol. 88, no. 3, pp. 44.
17. Xu Y., Xu H. and Zhang D. A novel disjoint community detection algorithm for social networks based on backbone degree and expansion, *Expert Systems with Applications*, 2015, Vol. 42, pp. 8349-8360.
18. Yang B., He H. and Hu X. Detecting community structure in networks via consensus dynamics and spatial transformation, *Physica A*, Vol. 483, pp. 156-170.
19. Bertsekas D.P. Constrained Optimization and Lagrange Multiplier Methods, Belmont, MA: Athena Scientific, 1996.
20. Feynman R., Leighton R. and Sands M. The Feynman lectures on physics, Vol. 1, Reading, Mass: Addison-Wesley Publishing Company Inc., 1963.
21. Horn R. u Johnson C. Matrix analysis, Second Edition ред., New York: Cambridge University Press, 2013.
22. Zachary W. An Information Flow Model for Conflict and Fission in Small Groups, *Journal of Anthropological Research*, 1977, Vol. 33, no. 4, pp. 452-473.

23. Lusseau D., Schneider K., Boisseau O., Haase P., Slooten E. and Dawson S. The bottlenose dolphin community of doubtful sound features a large proportion of long-lasting associations, *Behav. Ecol. Sociobiol.* 54 (4) (2003) 396–405, *Behav. Ecol. Sociobiol.*, 2003, Vol. 54 (4), pp. 396-405.
24. Lusseau D. and Newman M.E.J. Identifying the role that animals play in their social networks, in *Proceedings of the Royal Society of London. B, Biological Sciences*, 2004.
25. Girvan M. and Newman M.E.J. Community structure in social and biological networks, in *Proceedings of the National Academy of Sciences*, 2002.
26. Ferreira F., Jalali M. and Ferreira J. Integrating qualitative comparative analysis (QCA) and fuzzy cognitive maps (FCM) to enhance the selection of independent variables, *Journal of Business Research*, 2016, Vol. 69, pp. 1471-1478.
27. Newman M. Fast algorithm for detecting community structure in networks, *Physical Review E*, 2004, Vol. 69, pp. 1-5.
28. Wu Q., Qi X., Fuller E. and Zhang C.-Q. Follow the Leader”: A Centrality Guided Clustering and Its Application to Social Network Analysis, *Scientific World Journal*, 2013, Vol. pp. 1-9.
29. Newman M.E.J. Detecting community structure in networks, *The European Physical Journal B*, 2004, Vol. 38, no. 2, pp. 321-330.
30. Zhao P. and Zhang C.-Q. A new clustering method and its application in social networks," *Pattern Recognition Letters*, 2011, Vol. 32, pp. 2109-2118.
31. Balakrishnan H. and Deo N. Detecting Communities using Bibliographic Metrics, in *IEEE International Conference on Granular Computing*, 2006.
32. Fortunato S., Latora V. and Marchiori M. A Method to Find Community Structures Based on Information Centrality, 2004.
33. Lusseau D. The emergent properties of a dolphin social network, in *Proceedings of the Royal Society of London Series B-Biological Sciences*, 2003.

Статью рекомендовал к опубликованию д.т.н., профессор А.В. Боженюк.

Целых Александр Николаевич – Южный федеральный университет; e-mail: ant@sfedu.ru; 347928, г. Таганрог, пер. Некрасовский, 44; тел.: +79185562047; кафедра ИАСБ; д.т.н.; профессор.

Васильев Владислав Сергеевич – e-mail: vsvasilev@sfedu.ru; тел.: +79185983647; кафедра ИАСБ; к.т.н.; доцент.

Целых Лариса Анатольевна – Таганрогский институт имени А.П. Чехова (филиал) Ростовского государственного экономического университета (РИНХ); e-mail: l.telykh58@gmail.com; 347936, г. Таганрог, ул. Инициативная, 48; тел.: +79897207928; кафедра экономики и предпринимательства; к.э.н.; доцент.

Tselykh Alexander Nikolaevich – Southern Federal University»; e-mail: ant@sfedu.ru; 44, Nekrasovskiy, Taganrog, 347928, Russia; phone: +79185562047; the department IASB; dr. of eng. sc.; professor.

Vasilev Vladislav Sergeevich – e-mail: vsvasilev@sfedu.ru; phone: +79185983647; the department IASB; dr. of eng. sc.; the senior lecturer.

Tselykh Larisa Anatolievna – Chekhov Taganrog Institute (branch) of Rostov State University of Economics; e-mail: l.telykh58@gmail.com; Initsiativnaya, 48, Taganrog, 347936, Russia; phone: +79897207928; the department of Economics and business; dr. of ec. sc., the senior lecturer.