

---

---

## **СТАТИСТИЧЕСКИЕ МЕТОДЫ И АНАЛИЗ ДАННЫХ**

### **МЕТОДЫ АНАЛИЗА СОЦИАЛЬНЫХ СЕТЕЙ НА ПРИМЕРЕ ВИЗУАЛИЗАЦИИ СТРУКТУРЫ ПРЕДПОЧТЕНИЯ ПРОФЕССИЙ<sup>1</sup>**

В.В.Печенкин

*(Саратов)*

В статье продемонстрирована техника анализа социальных сетей на примере изучения престижности профессий. Описывается способ визуализации социальных сетей. Приводится пример применения этого алгоритма к данным прямого ранжирования профессий. Результат визуализации сравнивается с результатом построения иерархии доминирования.

*Ключевые слова:* социальная сеть, визуализация, визуальная деформация, модельное пространство, допустимая позиция, валидность, ранжирование.

#### *Анализ социальных сетей*

Анализ социальных сетей фокусируется на определении структуры взаимодействия социальных объектов. Можно выделить два подхода к использованию аппарата теории сетей<sup>2</sup> в этой области. В рамках первого подхода строится сетевая модель изу-

---

<sup>1</sup> Эмпирические данные получены в ходе пилотажного исследования проблемы преподавания информационных технологий в высших учебных заведениях города Саратова. Опрашивались студенты четырех вузов города (150 анкет).

<sup>2</sup> Существуют математические теории, которые оперируют понятием “сеть”. Нас интересует визуальное представление сетей, которое являет-

чаемой социальной структуры (предполагается, что репрезентативность и валидность этой модели обосновываются в процессе ее построения), затем описываются свойства, полученные математическими методами. Эти свойства интерпретируются исходя из позиций и концепций исследователя. Примером такого подхода является исследование [1], в котором для анализа социальной сети применяются алгоритмы поиска минимального остовного дерева и максимального потока.

Один из исследователей, активно развивающих второй подход, Линтон Фримен, так описывает основную идею сетевого анализа: «Ключевым фактором является роль визуального восприятия социальных сетей, которое позволяет исследователю по-новому открыть для себя социальную структуру и помогает передать это знание другим» [3]. Сетевой анализ в этом случае опирается на интуитивно понятную идею изучения образцов структурного взаимодействия индивидов и социальных групп в их социальной жизни. С позиции исследователя важным является то, что жизнь социального объекта зависит от того, как она вплетена в сети его социальных контактов. По одной из точек зрения [8], подтверждающих этот тезис, динамика развития социального объекта во многом зависит и от его структурных свойств.

В предлагаемой статье используется именно второй подход к исследованию социальных сетей. Их визуальное представление уже само по себе является результатом и источником аналитической информации. Кроме того, как будет показано, методы визуализации позволяют увидеть многомерность структуры изучаемых данных, которая затушевывается при построении одномерной шкалы измерения некоторого свойства.

---

ся одной из задач современной теории графов. Мы намеренно не обсуждаем понятия теории графов и подходов к математическому решению этой задачи. В данной работе нас интересует только визуальное восприятие информации о структуре данных, возможность интерпретации этой структуры.

Для описания социальных сетей используются два различных представления. Одно базируется на визуальном изображении с помощью вершин и соединяющих их линий, второе - на матричном представлении. В большинстве случаев при визуальном представлении вершинами являются объекты, линии (ребра, дуги) представляют связи, взаимодействие между этими единицами. В матричном представлении столбцы и строки, в свою очередь, представляют объекты, числа в ячейках матрицы характеризуют наличие взаимодействия и параметры взаимодействия этих объектов. Мы будем использовать оба упомянутых представления, помня о важности для нас именно визуального.

Типичный набор сетевых данных представляет собой два множества - вершин сети и отношений между этими вершинами. При этом можно представлять вершинами как индивидов (группы, организации), так и другие элементы социального пространства (например, статусные позиции). Отношения между вершинами отражают, например, наличие контактов между элементами, значимой корреляции между факторами в факторном пространстве и многое другое. Данные могут и не содержать информацию о реальных свойствах объектов, представляемых вершинами. Типичная проблема, которая решается в процессе визуализации, состоит в классификации и визуальном размещении в модельном пространстве вершин на основе их отношений с другими вершинами. Эти отношения могут описываться в модели направлением взаимодействия, подобностью структуры, расстояниями до других объектов, характеризующими их близость. Под модельным пространством мы понимаем множество допустимых точек, в которых можно размещать вершины сети. Эти точки называются допустимыми позициями. Модельное пространство определяет целый класс визуальных представлений сети и накладывает ограничения на способы манипулирования ее вершинами. Перемещение вершин сети из одной позиции модельного пространства в другую (визуальная деформация), таким образом, изменяет ее представление.

Обсуждая возможности визуализации социальных структур, которые позволят легче воспринять информацию о структурных свойствах данных, Фримен выделяет следующие моменты [3]:

- 1) Для лучшего восприятия свойств вершин используются вершины разного цвета и разного типа (окружности, прямоугольники, треугольники...).
- 2) Цвет и тип линий, соединяющих вершины, используются для показа различных типов отношений, связывающих элементы социальной сети.
- 3) Использование динамических визуальных эффектов позволяет представить изменения структуры сети во времени.
- 4) Визуальная воспринимаемость сетей может быть улучшена изменением взаимного расположения вершин, соответствующих социальным элементам.

Многие статистические процедуры, применяемые для обработки эмпирических данных в социологии, часто выдают большие объемы информации (нетрудно найти примеры, когда объем вторичных данных статистического анализа сопоставим с объемом первоначальной информации), которая, в свою очередь, требует исследования, увеличивает число возможных вариантов анализа. Общая стратегия при этом состоит в уменьшении количества анализируемой информации и разработке концепции, позволяющей учесть специфику изучаемой области. Необходимо заметить, что мы не предлагаем отказаться от статистически обоснованных процедур обработки данных. Аппарат визуализации социальных сетей дополняет статистические процедуры инструментарием описания структурных свойств изучаемых социальных объектов. Этот инструментарий полезен, когда он позволяет «ухватить» важные черты изучаемой информации. Степень простоты получаемых при этом моделей и их соответствие эмпирическим данным необходимо рассматривать в контексте самих данных. Кроме того, даже простая модель может оказаться полезной. Это зависит от того, как она представляет важные для исследо-

вателя свойства изучаемой структуры и насколько важны при этом потерянные характеристики.

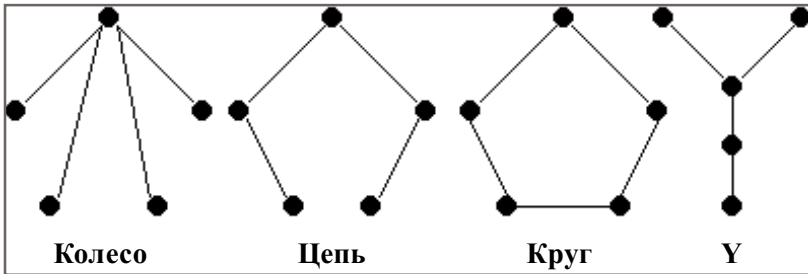
С самого начала сетевой подход к изучению поведения людей опирался на два принципа: (1) использование теории, основанной на математическом базисе, (2) систематический анализ эмпирических данных. Такой вид анализа социологических данных стал привлекательным только после того, как, начиная с 70-х годов, появились доступные в использовании, разнообразные компьютерные программы, оперирующие математическими объектами. Особенно это касается компьютерных программ, позволяющих работать с объектами комбинаторики и теории графов. Начиная с этого момента, анализ социальных сетей начал превращаться в междисциплинарное научное направление [3].

Можно привести много примеров применения такой техники. Одним из самых ранних примеров являются работы Морено, который в 30-х годах нашего века ввел в научный обиход идеи и инструменты социометрии [9, 10]. Позднее появились другие примеры использования инструментария сетевого социального анализа. В работе Левита<sup>1</sup> [7] исследовались эффективные способы социальных коммуникаций в группе. Он организовывал различные группы из 5 человек и давал определенное задание, которое позволяло оценивать эффективность структуры коммуникации в группе. Среди таких структур исследовались четыре - «Колесо», «Цепь», «Круг» и «Y» (см. рис. 1). Левит обнаружил, что структура «Колесо», при которой некий лидер занимает центральное положение, позволяет успешней выполнять задачи взаимодействия. При этом члены группы, которые находятся в центре системы коммуникации, были больше удовлетворены группой, периферийное же положение ограничивало самостоятельность.

Шепард [11] рассматривал разные типы данных как базис для многомерной процедуры шкалирования и предложил подхо-

---

<sup>1</sup>Краткое описание этой работы приведено в [12].



*Рис 1. Образцы структуры способов социальной коммуникации, исследуемых Левитом*

ды, которые позволяют выделить необходимую исследователю базовую информацию и отличать ее от технических преобразований, которые применяются к данным, и промежуточных результатов вычислений в процессе построения матрицы близости. Такая матрица может играть основную роль в процедуре визуализации социальной структуры.

Выбор способов определения измерительных характеристик для структур зависит от того, как эти характеристики будут преподнесены аудитории. Процедуры измерения обычно строятся так, чтобы результаты наилучшим образом соответствовали эмпирическим дистанциям в пространстве допустимых положений сетевой модели. Одна из таких процедур предложена Фрименом [4]. Она основана на концепции «степень близости к центру», которая использует в качестве основного критерия географическое расстояние и является достаточно информативной и интуитивно понятной.

Возможен вариант, когда для некоторых вершин нельзя найти точную позицию в модельном пространстве. Это связано со сложностью реальных данных и отсутствием алгоритмов наиболее общего решения задачи, которые за приемлемое время сформируют оптимальное размещение элементов. Один из таких алгоритмов описывается ниже.

### *Метод визуализации Кремпеля*

Рассмотрим метод визуализации социальных сетей, предложенный Лотаром Кремпелем [6]. Основное внимание в этом методе уделяется контролю валидности визуального представления структуры социального объекта, который подвергается упрощению и визуальной деформации. При этом выделяются принципы, которые необходимо принимать во внимание.

1. Выбор простых геометрических фигур, которые будут ограничивать допустимое пространственное размещение вершин сети<sup>1</sup> (вершины сети могут перемещаться только по границам этих фигур).
2. Выбор специальных характеристик структуры, которые можно использовать для оценки ее оптимальности (например, характеристики расстояния между социальными объектами).
3. Выбор эвристического алгоритма для оптимизации пространственного размещения элементов рассматриваемой структуры (под эвристическим в математике понимают алгоритм, который не гарантирует оптимального решения задачи, но позволяет получить его в приемлемые сроки).

Стратегия, предложенная Кремпелем, заключается в ограничении возможных позиций социальных объектов в процессе их визуализации простейшими образцами, о которых известно, что они легко воспринимаются аудиторией. В качестве примера можно рассматривать круг как пространственную фигуру для ограничения способов размещения элементов социальной сети. Существует, по крайней мере, одна формальная причина такого выбора. Если все элементы социальной сети равномерно размещены

---

<sup>1</sup> Вопрос размещения вершин сети решается социологами с использованием разных критериев. Например, Морено (см. [9, 10]) помещал вершины, соответствующие индивидам в позиции, отражающие их реальное положение в физическом пространстве.

на плоскости по окружности, то сумма геометрических расстояний от произвольного элемента до всех остальных одинакова. Этот способ, в частности, позволяет проиллюстрировать характер связей в сети и простым взглядом на визуальную модель увидеть специфические отклонения от идеального образца взаимодействия (например, нулевая гипотеза может состоять в том, что изучаемая структура соответствует этим идеализированным образцам). Конечно, мы не можем предложить универсального способа оценки величины отклонения от «идеальной» структуры, но в данный момент мы апеллируем именно к визуальной оценке и узнаванию структуры эмпирических данных (как это и происходит в упоминаемом выше исследовании Левита). Точно так же можно использовать другие фигуры - треугольники, квадраты, учитывая их специфические свойства.

Ограничение возможного местоположения в качестве решения задачи визуализации определяет и расстояния между допустимыми позициями вершин при построении представления. Если существует определенное число допустимых позиций в модельном пространстве, простая модель может быть описана матрицей расстояний между ними.

Алгоритм Кремпеля не предназначен для получения точного, с точки зрения выбранного критерия, варианта размещения вершин. Прагматическим подходом в этой ситуации является получение наиболее близкого к «точному» решения. При этом для визуальной оценки полученного решения зачастую достаточно разместить на оптимальных позициях лишь важнейшие элементы. Это также означает, что допускается наличие погрешности в размещении менее важных элементов структуры. Одна из главных идей предлагаемого алгоритма заключается в начальном размещении наиболее важных элементов, чтобы ограничить в дальнейшем число возможных позиций для менее важных элементов. Если наиболее важные элементы размещаются первыми, для оставшихся наименее важных элементов резко сокращается число допустимых позиции, что уменьшает вычислительную трудоемкость процедуры.



Шаги алгоритма, который может быть выполнен вручную или автоматизирован<sup>1</sup>, представлены ниже.

1. Отсортировать все элементы структуры (вершины сети) по их приоритету в соответствии с выбранным критерием сортировки.
2. Для всех вершин, начиная с наиболее важных, выполнить следующие шаги в сторону уменьшения их приоритета.
3. Оценить все допустимые позиции для вершин, которые связаны с текущей.
4. Поместить вершину в позицию пространства модели, которая еще не занята, и сумма расстояний до которой минимальна.
5. Считать позицию занятой, а элемент размещенным.
6. Повторить шаги (2-5) до тех пор, пока все элементы не будут размещены.
7. Повторить процедуру размещения (шаги 2-6) несколько раз до тех пор, пока визуальное представление не будет соответствовать субъективному критерию его оптимальности.

Для вершин, которые практически не могут быть различимы при их первичном упорядочивании, процедура должна назначать их начальный приоритет случайно в процессе повторяющихся процедур размещения. Описанная выше процедура использует понятия приоритета вершин и расстояния между вершинами, которые могут определяться исследователем самостоятельно, что позволяет применять ее для широкого спектра задач и интерпретаций.

---

<sup>1</sup> Существуют компьютерные программы, которые позволяют интерактивно редактировать визуальные структуры и используются в практике анализа социальных структур. Целый список таких программ доступен на сайте [http://www.heinz.cmu.edu/project/INSNA/soft\\_inf.html](http://www.heinz.cmu.edu/project/INSNA/soft_inf.html). Автор статьи использовал программу собственной разработки GRIN, описание которой и исполняемый файл программы (для Windows 9X, NT, 2000) располагаются по адресу [http://www.geocities.com/pechv\\_ru/](http://www.geocities.com/pechv_ru/).

### *Визуальное представление престижности профессий*

Применим алгоритм Кремпеля к эмпирическим данным о престижности профессий, которые были получены в ходе опроса студентов саратовских вузов в 2000 году по проблеме преподавания информационных технологий. Полученная далее сеть не является социальной сетью в привычном понимании [13], но в данном случае мы продемонстрируем применимость методов анализа к широкому классу задач визуализации.

Среди прочих вопросов в анкете предлагалось оценить престижность профессий по десятибалльной шкале. При этом список профессий был намеренно сокращен. В анкете были указаны только обобщенные названия, которые пользуются популярностью и достаточно часто звучат в средствах массовой информации (8 названий). Нас интересовало отношение студентов к профессиям, связанным с информационными технологиями (программист, администратор информационной системы), степень их престижности, соотношение субъективной оценки этих профессий с профессиями, популярность которых подтверждается конкурсом на соответствующие факультеты вузов (юрист, экономист). Методы изучения престижности профессий на основе их субъективных оценок подвергаются обоснованной критике, но продолжают использоваться, особенно это касается именно изучения престижности профессий [2].

Данные прямого ранжирования профессий преобразовывались в матрицу парных сравнений следующим образом. Была построена квадратная матрица  $A=(a_{ij})_{8 \times 8}$ , где каждой профессии соответствует одна строка и один столбец. Каждый элемент  $a_{ij}$  показывает, насколько чаще профессия  $i$  имела большую оценку в анкетах, чем профессия  $j$ .

Для упрощения матрицы  $A$  она была преобразована в матрицу  $B=(b_{ij})_{8 \times 8}$  по следующему правилу:

$$b_{i,j} = \begin{cases} 0, & \text{если } a_{i,j} \leq a_{j,i} \\ a_{i,j} - a_{j,i}, & \text{если } a_{i,j} > a_{j,i} \end{cases}$$

Итоговая матрица  $B$  показана в таблице 1. Например, число 46 в первом столбце таблицы означает, что число анкет, в которых профессия программиста получила более высокую оценку, чем профессия менеджера, минус число анкет с обратным соотношением этих же профессий, равно 46.

**Таблица 1.**

**ИТОГОВАЯ МАТРИЦА В ПРЕДПОЧТЕНИЯ ПРОФЕССИЙ**

	Мен-р	Бух-р	Аналитик	Прогр-ст	Юрист	Референт	Экон-ст	Адм. ИТ
Мен-р		20	8	0	0	47	0	0
Бух-р	0		0	0	0	0	0	0
Аналитик	0	18		0	0	0	0	0
Прогр-ст	46	44	47		0	39	0	0
Юрист	59	73	59	43		61	25	17
Референт	0	7	5	0	0		0	0
Экон-ст	23	47	24	10	0	55		0
Адм. ИТ	19	37	48	3	0	52	7	

Для визуализации данных, представленных в таблице, была построена сеть, вершинам которой соответствуют профессии, ориентированным дугам между вершинами соответствуют направления предпочтения профессий (если в таблице 1 значение элемента  $b_{ij}$  больше нуля, вершина  $i$  соединяется с вершиной  $j$  направленной дугой, которая помечена числом  $b_{ij}$ ). В качестве значений, по которым была проведена первичная сортировка, необходимая при применении алгоритма Кремпеля, были использованы максимальные значения в строке таблицы соответствующей профессии. При построении визуального изображения была реализована следующая стратегия размещения вершин: если имеется дуга от первой вершины ко второй, то вторая вершина размещается по окружности против часовой стрелки относительно

первой. Ближайшей вершиной к текущей считается та вершина, к которой ведет дуга с минимальным весом. После обработки полученной визуальной структуры в соответствии с алгоритмом было получено изображение, которое показывает структуру престижности профессий (рис. 2).

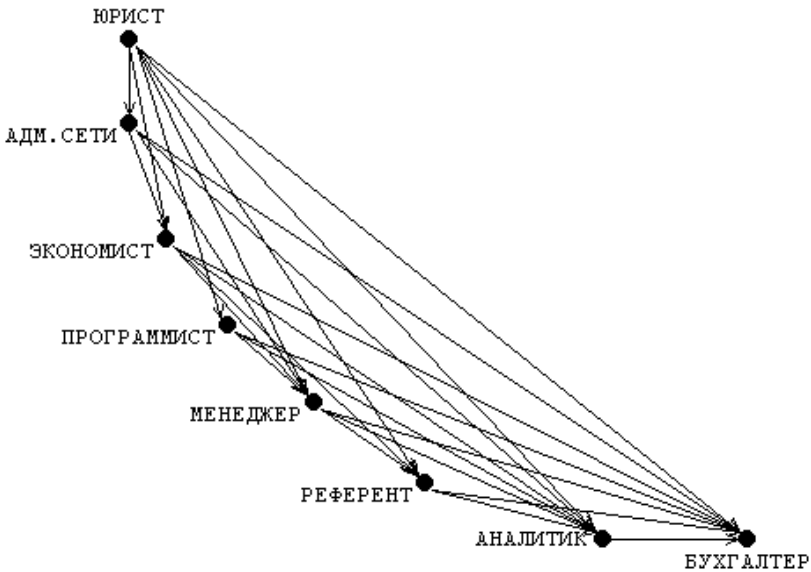


Рис. 2. Результат обработки информации алгоритмом Крмпеля

### *Упрощение и визуальная деформация представления*

Полученное изображение является достаточно сложным для восприятия и нуждается в дальнейшем упрощении. Большое количество дуг не позволяет понять специфику изучаемой структуры. Были удалены некоторые дуги, которые, на наш взгляд, содержали информацию, дублируемую другими элементами модели. В качестве нового модельного пространства была использована прямоугольная сетка на плоскости. В соответствии с новой стратегией визуальной деформации направления уменьшения

престижа профессии при перемещении вершин - вправо и вниз от текущей позиции вершины. Окончательный результат работы алгоритма Кремпеля с новым модельным пространством и новой стратегией перемещения вершин показан на рис. 3.



Рис. 3. Визуальное представление данных о предпочтении профессий

Получено не одномерное вертикальное ранжирование, а более сложная структура, которая показывает престижность профессий более рельефно. В данном случае в процессе упрощения сети мы смогли бы дойти до одномерной иерархии<sup>1</sup>, порядок расположения элементов в которой виден уже на рис. 2. Но дуги, оставленные в окончательном представлении, показывают многомерность изучаемой структуры. Числа на дугах (вес) показывают степень близости профессий в модельном пространстве.

Чтобы оценить новое визуальное представление, профессии были упорядочены методом, применяемым для построения иерархии доминирования на основе данных парного сравнения. Суть

<sup>1</sup>Мы не сможем получить одномерной иерархии престижности профессий, если сеть будет содержать циклы. Но и в этом случае мы можем предложить процедуры упрощения визуального представления, например, разрывая звено минимального веса любого цикла.

метода изложена в [5]. Результаты рассчитываются по следующей формуле:

$$S(Q_i) = \frac{2(W_i - L_i)}{N_i} + Q_i, \text{ где } S(Q_i) - \text{ оценка положения } i\text{-го объекта в иерархии, } L_i - \text{ число сравнений, в которых } i\text{-ый объект имеет меньшую оценку, } Q_i - \text{ среднее число оценок всех объектов, которые сравнивались с } i\text{-ым, } N_i - \text{ число сравнений } i\text{-го объекта с другими, } W_i - \text{ число сравнений, в которых } i\text{-ый объект имеет более высокую оценку. Оценки рекурсивно вычисляются по приведенной формуле до наступления стабилизации и достижения определенной точности.}$$

Результаты упорядочивания профессий приведены в таблице 2, где каждой из них соответствует ее позиция на одномерной шкале: от максимального значения 0,712 для профессии «Юрист» до минимального -0,5238 для профессии «Бухгалтер».

**Таблица 2.**

**РЕЗУЛЬТАТЫ УПОРЯДОЧИВАНИЯ ПРОФЕССИЙ**

Юрист	Адм. ИТ	Экон-ст	Прог-ст	Менеджер	Аналитик	Референт	Бух-р
0,712	0,3168	0,2655	0,2498	-0,1558	-0,378	-0,5229	-0,5238

Результаты визуализации согласуются с методом построения иерархического доминирования, но предлагают больше информации для интерпретации, позволяют варьировать степень упрощения структуры от самой полной информации о предпочтениях (например, результаты работы метода Кремпеля) до, практически, одномерной иерархии, которая в нашем случае соответствует результатам, приведенным в таблице 2.

Существует некоторое отличие в результатах ранжирования, полученных с помощью этих подходов. При построении иерархии доминирования, результаты которой приведены в таблице 2, профессия «Аналитик» престижней профессии «Референт», что отличается от результата визуализации на рис. 3. Мы можем объяс-

нить это расстоянием в модельном пространстве между этими профессиями и профессией «Менеджер». Это расстояние и предопределило положение двух профессий на шкале престижности. Но в данных парного сравнения (таблица 1) мы имеем более высокий престиж профессии «Референт» по отношению к профессии «Аналитик». Этот факт заставляет нас при отсутствии циклов в сети предпочтений поместить первую на более высокую позицию, как это сделано на рис. 3.

В данной работе мы не останавливаемся на интерпретации результатов визуального представления. Мы только продемонстрировали технику обработки эмпирической информации одним из методов анализа социальных сетей.

#### ЛИТЕРАТУРА

1. Градосельская Г.В. Социальные сети: обмен частными трансфертами// Социологический журнал, 1999. №2. С. 156-163.
2. Ильин В. Государство и социальная стратификация советского и постсоветского обществ. 1917-1996 гг.: Опыт конструктивистско - структуралистского анализа. Сыктывкар:Сыктывкарский университет, Институт социологии РАН, 1996.
3. Freeman L.C. Visualizing Social Networks. Journal of Social Structure, 2000. Vol. 1(1); или <http://www.heinz.cmu.edu/project/INSNA/joss/vsn.html>
4. Freeman L.C. Centrality in Social Networks: I. Conceptual clarification// Social Networks, 1979. №1. P. 215-239.
5. Appleby C., Jameson K.A., Freeman L. Finding an appropriate order for hierarchy based on probabilistic dominance. Animal behaviour, The association for Study of Animal Behaviour, 1999. Vol. 57. P. 991-998; или <http://www.idealibrary.com>.
6. Krempel L. Simple Representations of Complex Networks: Strategies for Visualizing Network Structure//[http://www.mpi-fg-koeln.mpg.de/~lk/algo5a/ algo5a.html](http://www.mpi-fg-koeln.mpg.de/~lk/algo5a/algo5a.html).
7. Leavitt J. Some effects of certain communication patterns on group performance//Journal of Abnormal and Social Psychology, 1951. Vol. 46. P. 38-50.
8. Mesarovic M.D., McGinnis D.L., West D.A. Cybernetics of Global Change: Human dimension and Managing of Complexity. MOST Policy Papers 3, 1996.

9. *Moreno J.L.* Application of the Group Method to Classification. New York:National Committee on Prisons and Prison Labor, 1932.
10. *Moreno J.L.* Who Shall Survive? Washington, DC:Nervous and Mental Disease Publishing Company, 1934.
11. *Shepard R.N.* A Taxonomy of Some Principal Types of Data and of Multi-dimensional Methods for their Analysis//Multidimensional Scaling, 1972. P.23-44.
12. *Смелзер Н.* Социология. М.:Феникс, 1998.
13. *Чураков А.Н.* Анализ социальных сетей//Социологические исследования, 2001. №1. С. 109-121.