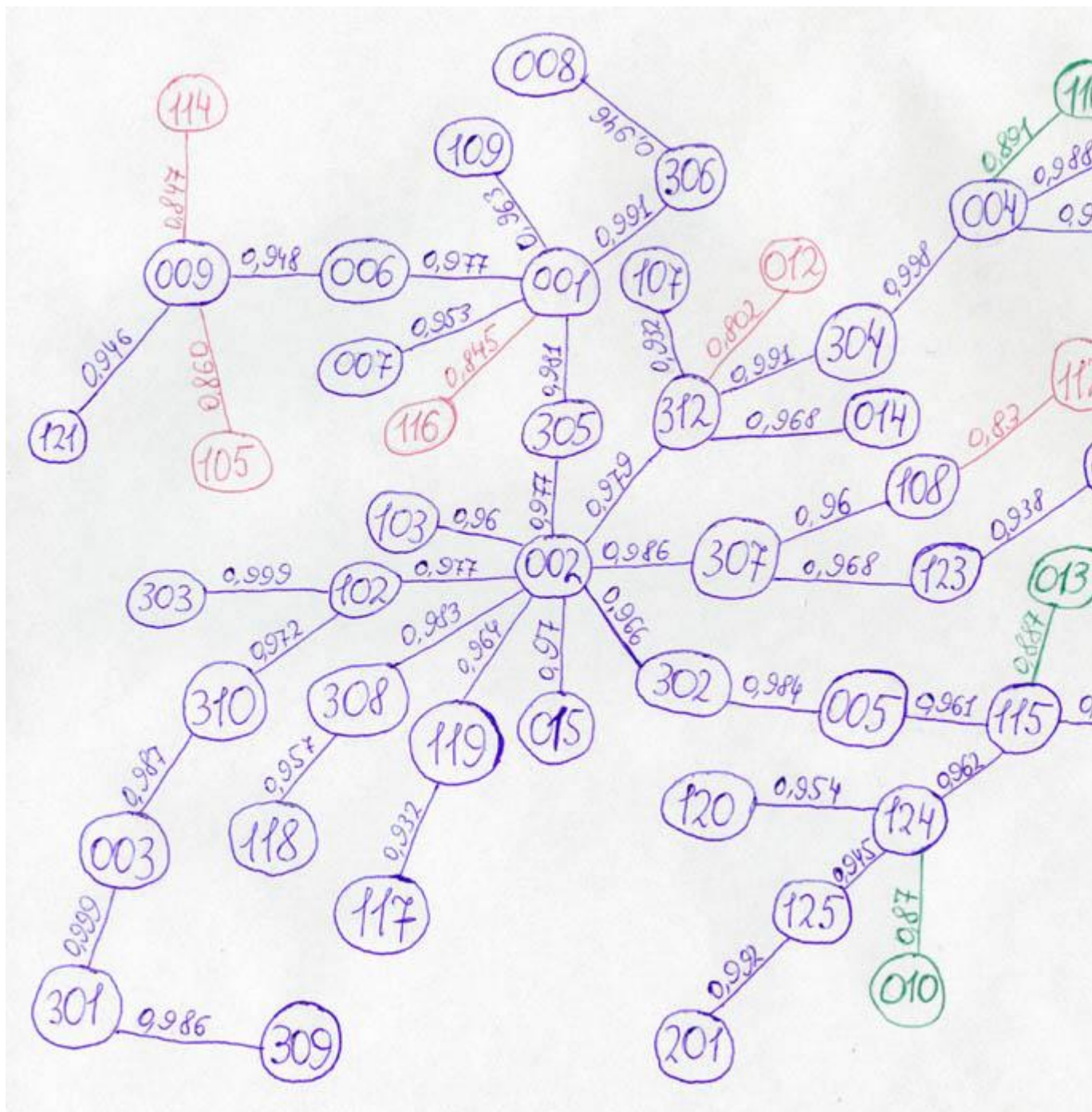


Корреляционные плеяды

Все чаще и чаще от студентов требуют в курсовых и дипломных работах строить корреляционные плеяды (корреляционные графы). Плеяды становятся неотъемлемой частью корреляционного исследования, необходимым условием того, что работа будет благосклонно принята научным руководителем и успешно пройдет защиту.

Данная тенденция, очевидно, связана с тем, что сейчас не составляет особого труда рассчитывать огромные корреляционные матрицы, отвечающие за взаимосвязи между исследуемыми параметрами. Множество найденных корреляций не всегда легко описать, сложно выделить важное и отбросить лишнее. Зачастую подсчитав корреляции, сложно понять с чего начать, собственно, анализ результатов. И тут, как нельзя, кстати, пришлись корреляционные плеяды, которые позволяют графически отобразить все многообразие взаимосвязей. О плеядах вспомнили те преподаватели и профессора, которые выросли на них, для которых они являются привычными. Они используют корреляционные плеяды в своих научных работах, зачастую вручную рисуя их на бумажке ручкой или карандашом, как это делали еще полвека назад.

Пример таких плеяд, можно увидеть, например, тут:



По сути, корреляционные плеяды – это способ отображения информации о корреляциях, который помогает их структурировать, проводить объединение коррелирующих факторов. Здесь стоит сделать небольшое отступление, и рассказать об учебном процессе: еще десятилетие назад никакой такой сложной статистики не требовали от студентов. С начала появления компьютеров в широком использовании, с ростом возможностей текстовых редакторов и электронных таблиц, студенту вполне хватало сравнительного анализа табличных данных, а до компьютерной эры не требовалось и того – достаточно было одного лишь описания частных случаев, выявления тенденций «на глазок».

История вопроса

Использовать корреляционные плеяды предложил в конце 50-х годов прошлого века ленинградский зоолог проф. П. В. Терентьев. Занимаясь изучением корреляций между множеством различных признаков озерной лягушки, он объединил их в группы по абсолютной величине коэффициентов корреляции, тем самым получив два распределения: признаки с малой и большой по величине корреляцией. Терентьев назвал эти группы корреляционными плеядами и опубликовал несколько способов их анализа.

Вслед за биологами новый метод быстро переняли психологи университета, и метод корреляционных плеяд стал одним из основных в дифференциальной психологии и многих других психологических науках.

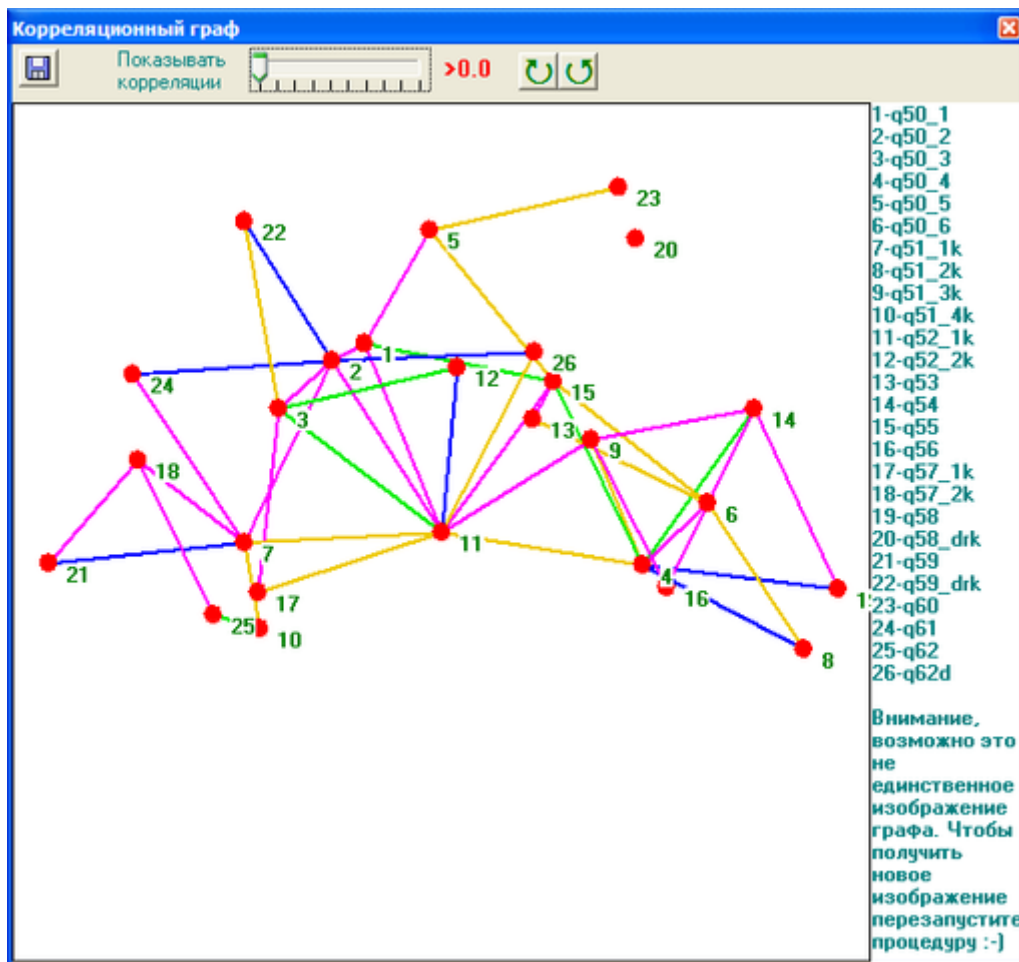
Методика корреляционных плеяд в дальнейшем была расширена Ростовской. [Ростова Н.С. Корреляции: структура и изменчивость. - СПб.: Изд-во С-Петербур. ун-та, - 2002, - 308с.].

К сожалению, современные технические возможности и доступность литературы, мало способствовали повышению общего уровня статистической грамотности среди студентов, как, впрочем, и остальных научных работников, включая, разумеется, и педагогический состав вузов. Перечислять все наиболее часто встречающиеся в работах ошибки и неточности – тема для отдельной статьи, которую, по-хорошему, наверное, следовало написать первой. Однако остановимся сейчас на помощи студентам в построении корреляционных плеяд. Тем более, что задача эта представляет, скорее даже не научную значимость, а почти непреодолимую техническую проблему для студента. Даже если студент понимает теоретическую основу построения плеяд, то грамотная их визуализация, без применения специальных программных средств, практически недостижима (за то время, которое есть в наличии у студента).

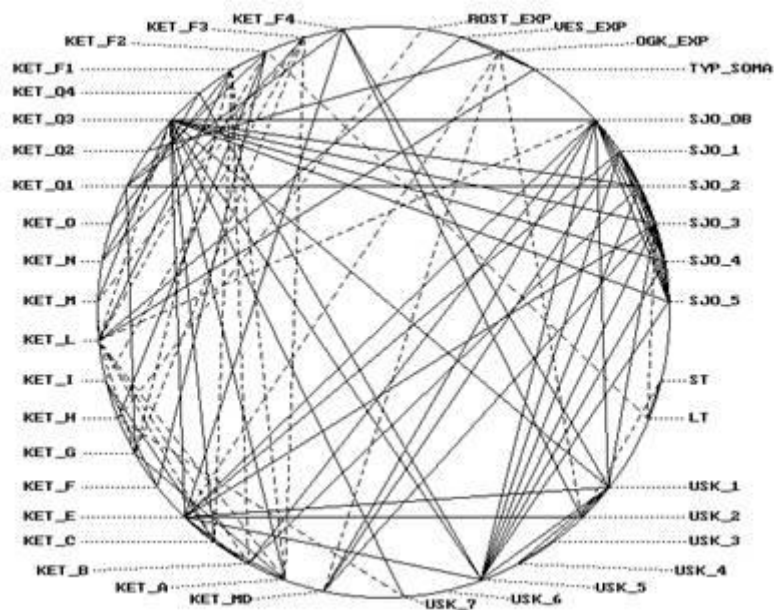
Как уже было сказано выше, метод корреляционных плеяд – это отечественная разработка. За полвека ее существования – она так и осталась внутренней методикой, никем практически вне России не используемой. Если мы посмотрим множество всех англоязычных работ по психологии, где будет встречаться корреляционная обработка, то ни в одной из них мы не встретим корреляционных плеяд (если, конечно, авторы не наши соотечественники). Для проверки этого факта, достаточно зайти на <http://scholar.google.com> и дать соответствующий запрос. К слову сказать, ни в одном англо-русском словаре нет сочетания "correlation pleiades".

Ни существует и ни одного программного продукта по статистике, в который изначально была бы встроена функция построения корреляционных плеяд. Это касается и двух самых мощных пакетов статистической обработки, существующих на сегодняшний момент, Statistica 8 и SPSS 17. Впрочем, практически в каждом мощном средстве статистической обработки есть встроенный язык программирования, и многие энтузиасты пытались реализовать подобную функцию самостоятельно, либо на основе уже существующих пакетов, либо с нуля. К сожалению, необходимость разбираться в сложных настройках и программных кодах, лишает обычных пользователей возможности пользоваться этими наработками. Далеко не всегда, кстати, хорошо обстоят дела с удобочитаемостью, получаемых таким образом графиков.

Вот пример:



Или вот:



Попробуйте тут разобраться, что с чем коррелирует.

Корреляционные плеяды – а нужны ли они вообще?

Существует мнение, с нашей точки зрения вполне оправданное, что в построении корреляционных плеяд скрыто больше отрицательных, чем положительных моментов. Основным минусом является то, что, по сути, корреляционная плеяда представляет собой набор парных корреляций, изображенных так, будто речь идет о взаимосвязи нескольких переменных. Это может натолкнуть неопытного исследователя на ложные выводы, что, надо отметить, в условиях российского образования, не сделает его работу менее защищаемой, а скорее даже наоборот: избытие любых, пусть и неоправданных, заключений отчего-то всегда у нас приветствуется. Все что можно вытянуть и что подразумевает под собой корреляционная плеяда, дает факторный анализ. Конечно, в отсутствии компьютера – вопросов нет, плеяды нарисовать проще, чем просчитать на калькуляторе факторный анализ, но когда факторный анализ можно провести за несколько кликов мышки, то нужда в изображении плеяд у исследователя пропадает. Современные средства анализа и визуализации статистических данных много глубже и точнее отображают реальность, чем корреляционные плеяды, на которых можно зачастую «увидеть» то, чего на самом деле нет.

Сочувствующие корреляционным плеядам ученые, разбирающиеся в статистике, полагают, что плеяды все же можно использовать, как подспорье последующего моделирования структурными уравнениями. Но в рамках дипломной работы, студенту уже плохо и от самих плеяд, не то что от структурных уравнений.

Строим, как можем :)

Вообще-то весь этот разговор о ненужности корреляционных плеяд не имеет особого смысла, потому что реальность такова, что чтобы сдать дипломную или курсовую работу студенту надо построить корреляционные плеяды. Рассмотрим, как это обычно происходит. Вот стандартный пример на построение плеяд, с которым студенты приходят в нашу компанию, в рамках написания своих учебных работ.

Задача обычно ставится так: студент провел исследование, опросил 10-100 человек по 2-7 многофакторным методикам, подсчитал ключи к ним, потом где-нибудь на кафедре, по этим ключам, ему подсчитали корреляционную матрицу. При этом любопытным фактом является то, что почти у каждой кафедры, у каждого преподавателя существует свое предпочтение относительно выбора программного средства: SPSS или Statistica. Результаты, полученные в «ненадлежащем средстве», не принимаются и не рассматриваются. Противостояние чем-то похоже на противостояние между московской и ленинградской школами марксизма-ленинизма, с той только разницей, что результаты в обеих программах получаются абсолютно одинаковыми, а теории марксизма-ленинизма в Москве и Ленинграде действительно несколько различались.

Итак, пусть будет 5 методик по 5 факторов в каждой. Следовательно, всего исследуется 25 факторов. Корреляционная матрица будет представлять из себя квадрат размером 25x25. Всего 625 ячеек. Тут, к слову сказать, что далеко не каждая из этих ячеек нужна для дальнейшего анализа: во-первых, матрица симметричная, что вдвое сокращает необходимое число, а, во-вторых, нас не интересует центральная диагональ (корреляция фактора самого с собой). Из элементарных основ комбинаторики подсчитаем: попарное сравнение 25 факторов даст нам $25 \cdot 24 / 2 = 300$ сравнений. 300 пар, для которых подсчитан коэффициент корреляции. В реальных исследованиях число коррелирующих факторов

может достигать хоть 100%. Обычно, на 300 сравниваемых пар где-то в среднем найдется порядка 30-60 для которых уровень корреляций будет признан значимым. При этом совсем некоррелирующих ни с кем другим факторов обычно не бывает. Итого, в нашей корреляционной плеяде должны участвовать порядка 20 факторов, суммарно между которыми будет порядка 40 взаимосвязей. Это граф с 20 вершинами и с 40 ребрами. При этом, скорее всего, этот граф не будет связным. Иначе говоря, наша корреляционная плеяда будет представлять из себя несколько связных подграфов («корреляционных островов») – маленьких корреляционных плеяд, не имеющих связей-мостов между друг другом.

Увидеть какой будет корреляционная плеяда, сколько в ней будет компонентов связности, глядя на исходную корреляционную матрицу совершенно невозможно, если вы не «человек дождя».

Корреляционная матрица – это худшая отправная точка для построения корреляционных плеяд. Во-первых, она содержит много излишней информации, во-вторых, выудить из нее нужную информацию в требуемом виде – непростая задача. Хорошо еще, если студент заранее посмотрел, какой порог должен преодолевать коэффициент корреляции, при выбранном им уровне значимости. Если есть возможность, то в программе статистической обработки следует в качестве вывода указывать не корреляционную матрицу, а столбец попарных корреляционных взаимосвязей (к сожалению, далеко не все программы статистической обработки позволяют делать это). Если удастся получить такой столбец, то в программе MS Excel можно в два счета отсортировать корреляции в порядке убывания тесноты связи (не забывая что убывать должна абсолютная величина (величина, взятая по модулю)). Такой отранжированный столбец – и есть правильная отправная точка для построения плеяд.

Если столбца нет, но есть матрица – что ж, придется повозиться. Из матрицы путем нескольких нехитрых преобразований в MS Excel можно получить тот же самый столбец. Т.к. задачу построения плеяд студенту решать приходится нечасто, то сойдет совсем уж бесхитростный путь в лоб – можно разрезать матрицы на столбцы, выставить их друг под другом, при этом дополнив названия сравниваемых параметров слева. Когда вся матрица будет порезана – удалим дубли и отсортируем по модулю коэффициенты корреляции. Имея такой столбец, исследователь, по крайней мере, не потеряет значимые корреляции, а более того в первую очередь будет обращать внимание на пары, связь между которыми наиболее сильная.

Теперь рассмотрим, что обычно получается у студентов разной степени добросовестности, пытающихся построить корреляционные плеяды.

Как мы подсчитали выше, корреляционная плеяда в среднем – это граф с множеством вершин и с множеством связей. Пытаться изобразить такой граф без предварительной подготовки, означает сделать его либо неудобочитаемым, либо слишком сильно упростить, потеряв важную информацию.

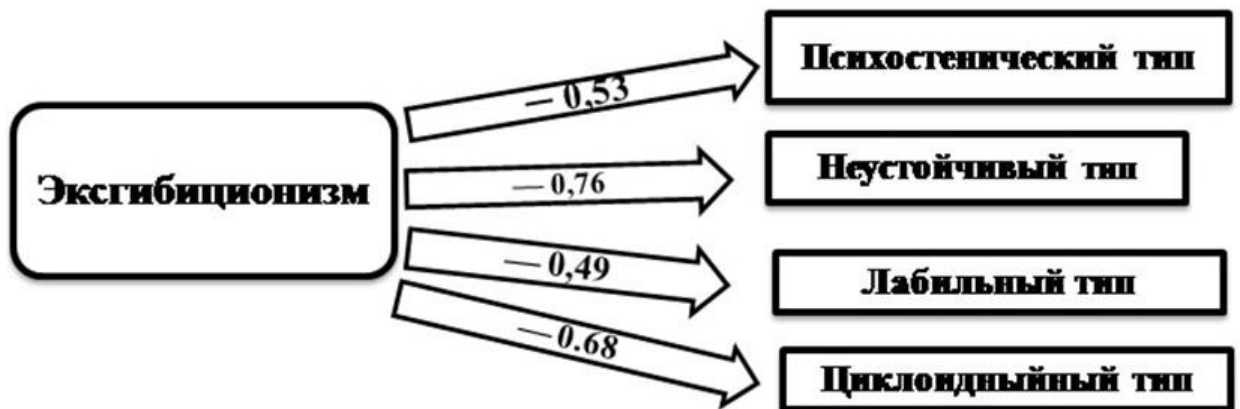
Не найдя подходящего программного решения по изображению плеяд, многие приходят к выводу, что для отображения корреляционных плеяд – самое лучшее – это обычный

графический редактор. Но попробуйте в графическом редакторе или при помощи фигур Word'a изобразить достаточно сложный граф. Тут-то и начинают студентам (и не только им) приходить на ум идеи оптимизации такого рисования. Этим идеям условно две: сокращение и извращение. Часто эти идеи применяются одновременно.

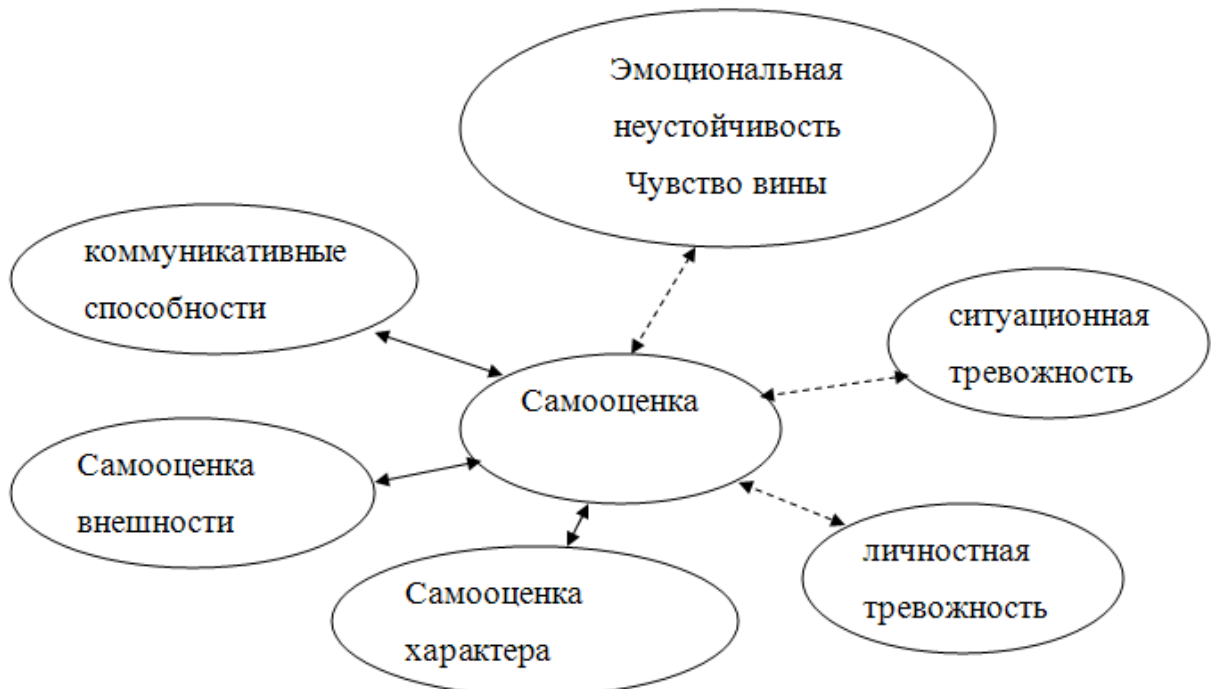
Наивное упрощение (выкидывание большей части связей)

Название этого примера отражает его суть. Студент выбрасывает из построения большинство корреляций, указывая лишь те, которые он нашел первыми, либо те, которые он может легко объяснить.

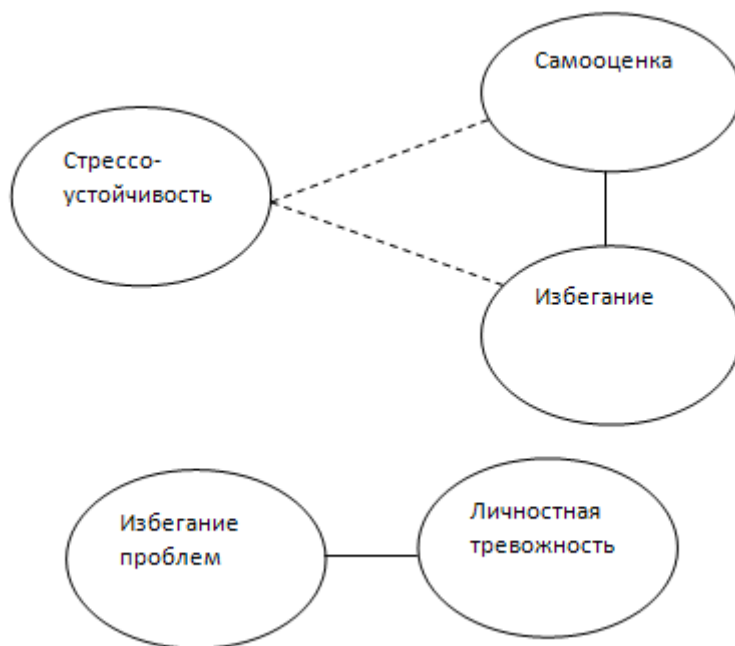
Пример 1. Указывание корреляций только для одного фактора:



Или:



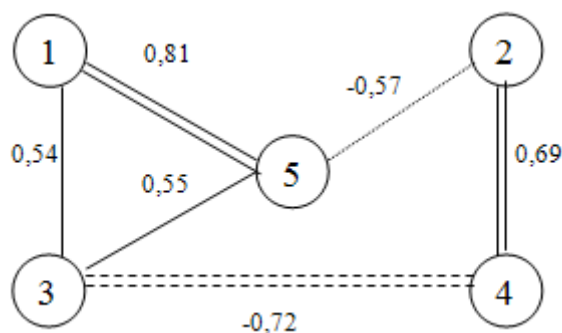
Указывание корреляций для нескольких факторов (но указываются далеко не все корреляции).



Как мы видим, во всех вышеуказанных попытках, студенты сохранили название факторов внутри вершин графа. Этот, заслуживающий уважения подход, приводит к хорошей «читаемости» плеяд, но по сути – это не корректные плеяды, т.к. вместо полной картины, мы видим лишь небольшие фрагменты. Суть плеяд от такого построения в значительной степени теряется.

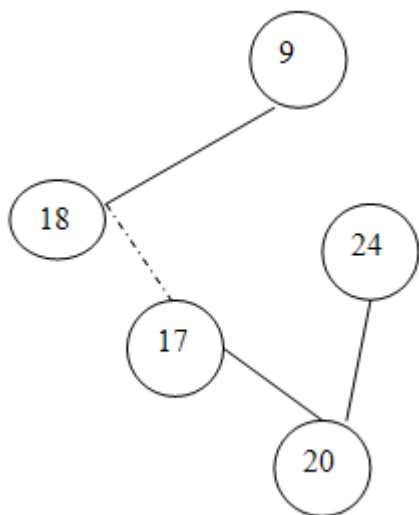
«Идеальные» примеры, из которых ничего не понятно

Несколько иной подход предлагается в методических указаниях различных вузов. Вот такой пример приведен в методичке Санкт-Петербургского института психологии и акмеологии:



Прямые корреляции принято изображать сплошной чертой; обратные – пунктиром, сообщается в методичке.

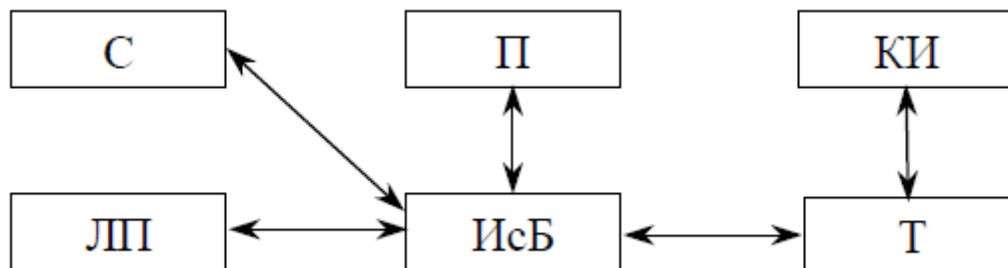
Как видно из примера, проблема с местом на листе теперь решена. Название каждого фактора заменено уникальным числовым индексом. И, похоже, что все корреляционные взаимосвязи теперь указаны для каждого фактора: т.е. упрощения структуры нет. Но, во-первых, сложно воспринимать такой граф: приходится глядеть, то на картинку с индексами, то на расшифровку этих индексов, а во-вторых – какой-то уж больно простой случай рассматривается в методичке. Всего 5 факторов и 6 взаимосвязей. Не удивительно, что



Тут потеряно почти все. При таком упрощении, вместо индексов можно было бы использовать и названия факторов: очень сложно не поместить на экран пять кружочков со словами внутри :)

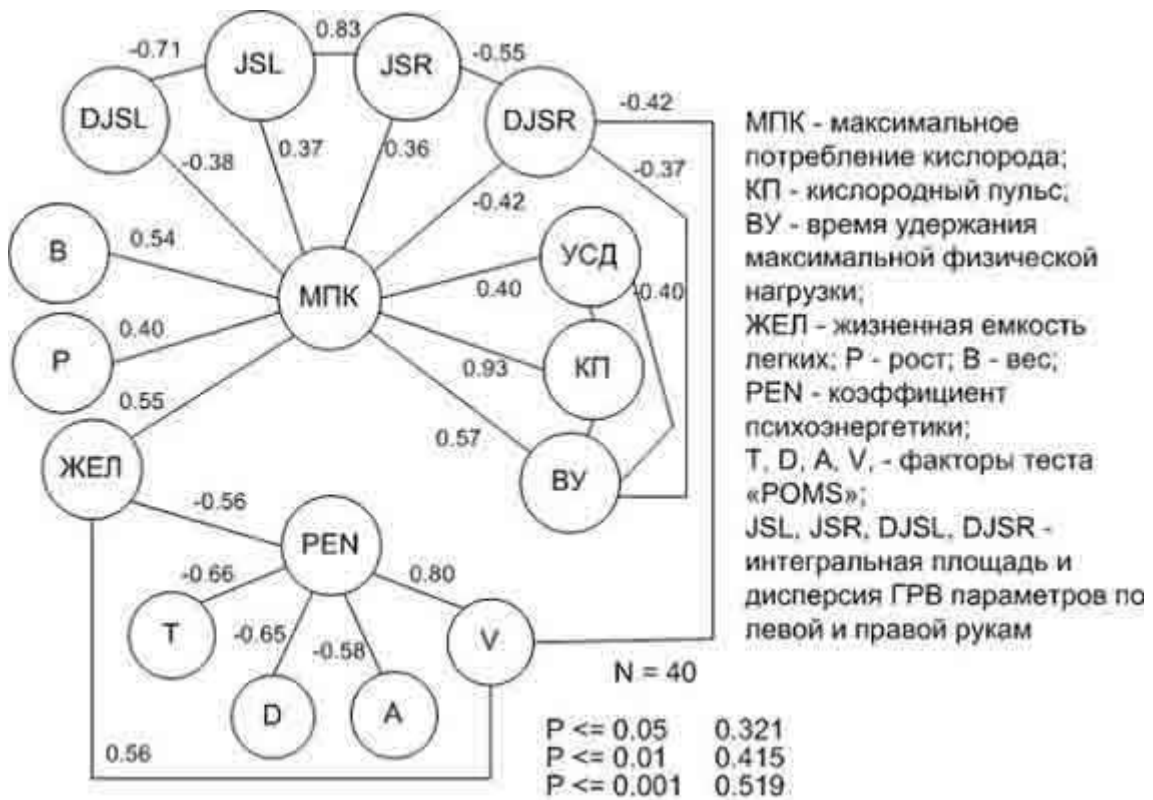
Буквенное кодирование

Вместо числовых индексов удачной идеей является использование буквенного кодирования факторов. Человек в теме, быстро сообразит, какие буквы за какой фактор отвечают, и ему не придется постоянно переключать свое внимание.



Вот добротное исполнение плеяды, взятое из одной диссертации.

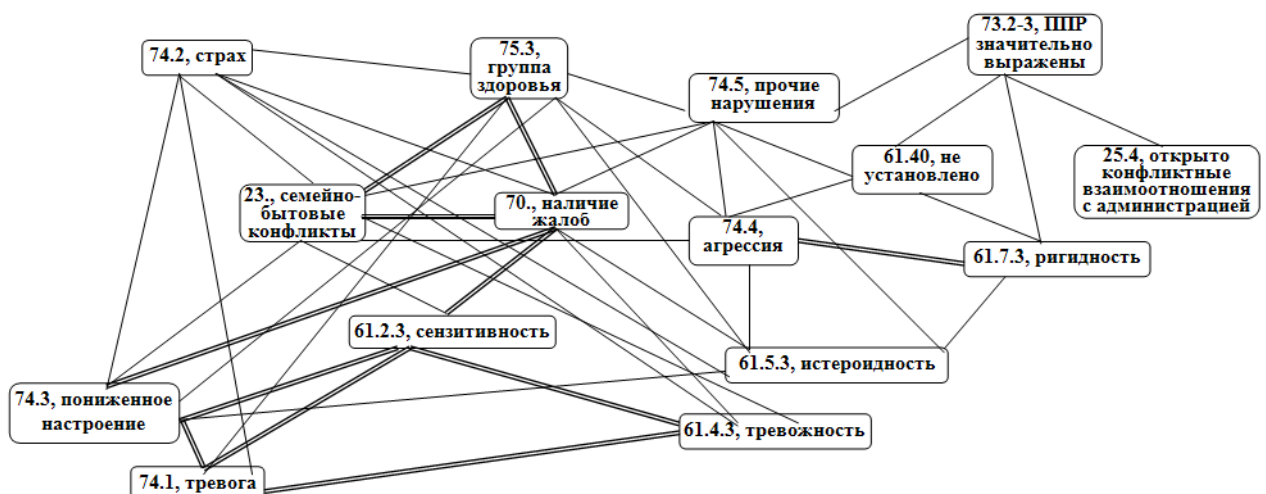
Еще одно удачное построение плеяды с использованием, очевидно, специальных программных средств для визуализации графа:



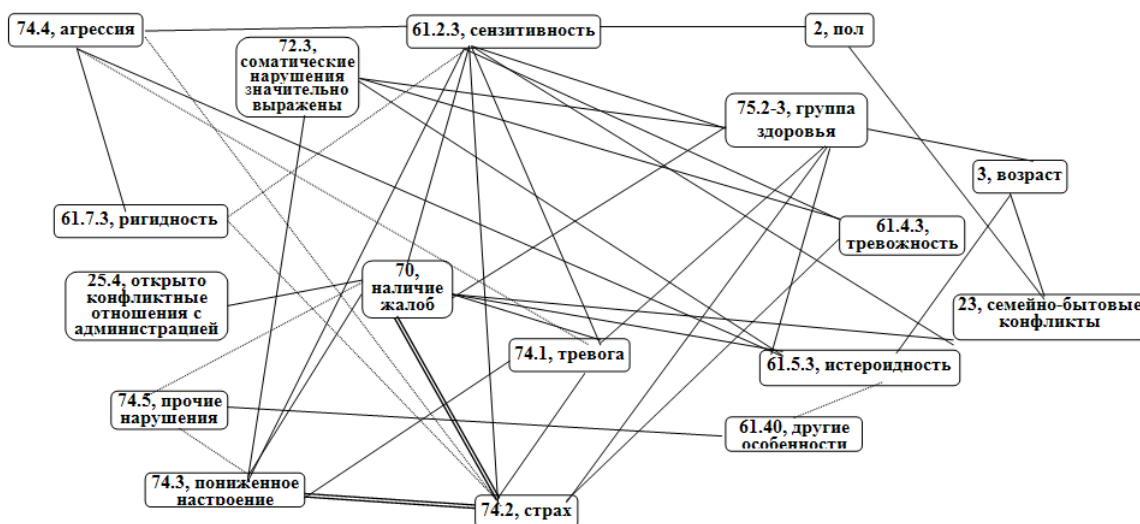
Без упрощений и сокращений: подход сильных духом

Все вышеперечисленные способы визуализации корреляционных плеяд не лишены недостатков. Однако нельзя пройти мимо по истине грандиозной работы тех добросовестных людей, которым важно сохранить все плюсы от словесного отображения факторов, но которые, при этом, не идут ни на какую сделку с совестью – не выкидывают из графика ни единой корреляционной взаимосвязи.

Так выглядит обычно результат их титанического труда:



Или так:



Если вы хотите сделать приятное ученому совету, или своему научному руководителю, то можете смело начать рисовать нечто подобное :)

На основе вышеперечисленного, можно сделать лишь один вывод: практически любая попытка визуализировать корреляционную плеяду ведет либо к потере возможности ее интерпретировать, либо к ее значительному упрощению. Таким образом, смысл корреляционных плеяд, как средства визуализации корреляций с целью упрощения их восприятия – теряется: либо мы видим не все корреляции, либо мы не понимаем, что, с чем коррелирует.

Несколько советов в конце :)

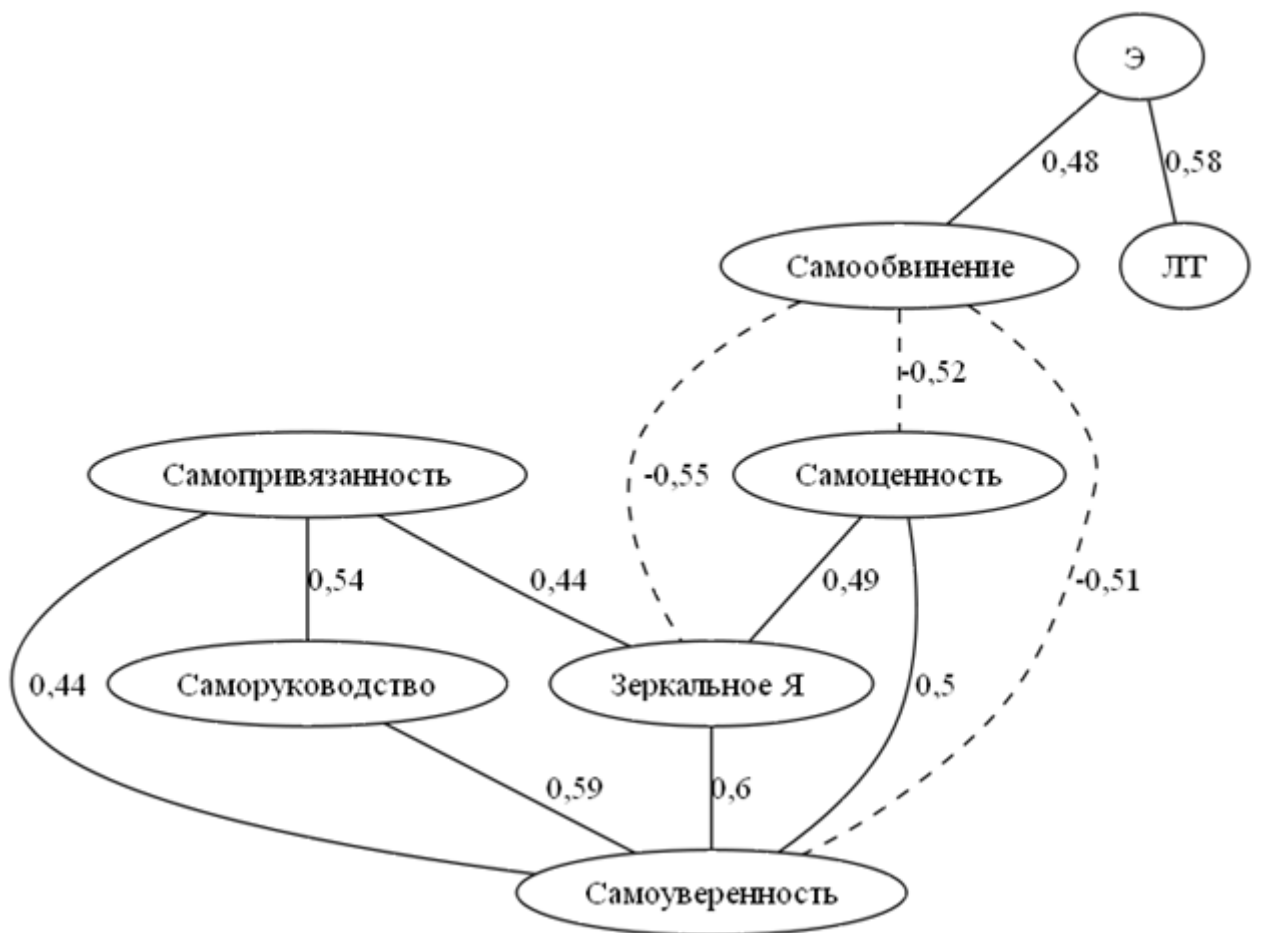
Чтобы не заканчивать на грустной ноте, дадим несколько советов как можно все же упростить построение плеяд и сделать их доступными для понимания.

Выше было описано, каким образом можно получить столбец коэффициентов корреляции для всех пар факторов, где эти коэффициенты признаны значимыми.

Теперь дело за малым. Надо найти программу, которая по этому столбцу построит необходимый график. Такие программы есть, к примеру, известный бесплатный продукт для визуализации графов Gepviz. При помощи нескольких преобразований в Excel'е можно сделать из столбца корреляций между факторами – входной файл для программы визуализации.

При этом, в зависимости от выбранных алгоритмов отображения графа, программа сама представит корреляционную плеяду на плоскости наилучшим образом: например так, чтобы она была наиболее компактна и чтобы ребра графа имели наименьшее число пересечений. Если корреляционная плеяда состоит из нескольких компонент связности, то программа изобразит сразу их все – каждую в своем компактном виде.

Пример графика:



Коллектив авторов Роботрона :)