Статистическое моделирование сетевых структур: теория, методология и практика

София Докука

Москва, 2017

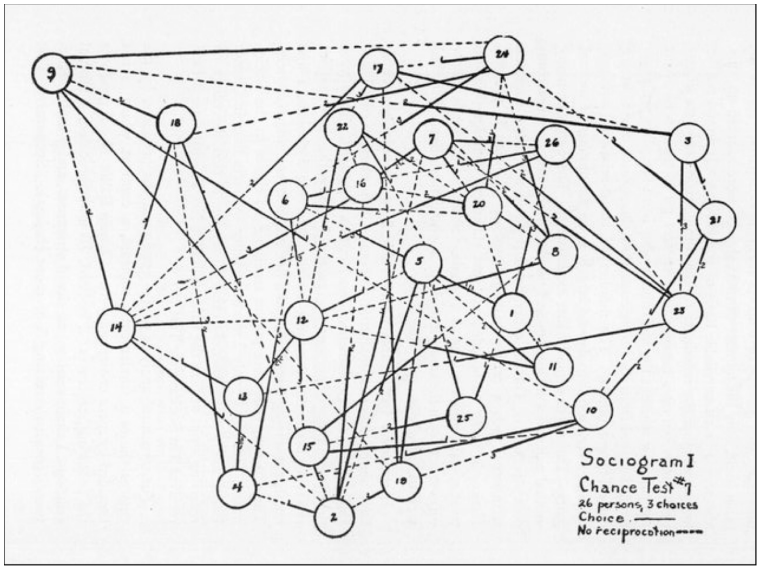
Сетевой подход

Анализ сетей различной природы широко используется в различных теоретических и эмпирических исследованиях.

- Биология и медицина
- Физика
- Прикладная математика и компьютерные науки
- Общественные науки: **социология**, экономика, политология и антропология

Анализ социальных сетей

Родоначальником анализа **социальных** сетей принято считать психолога Якоба Морено, который в 1930-х годах изучал влияние социального окружения на благополучие индивидов.
В те годы была разработана методология сбора данных в социальных сетях с помощью опросов (*социометрический метод сбора данных*).



Морено с коллегами издавал журнал **Sociometrics**.

Развитие методологии сетевых исследований

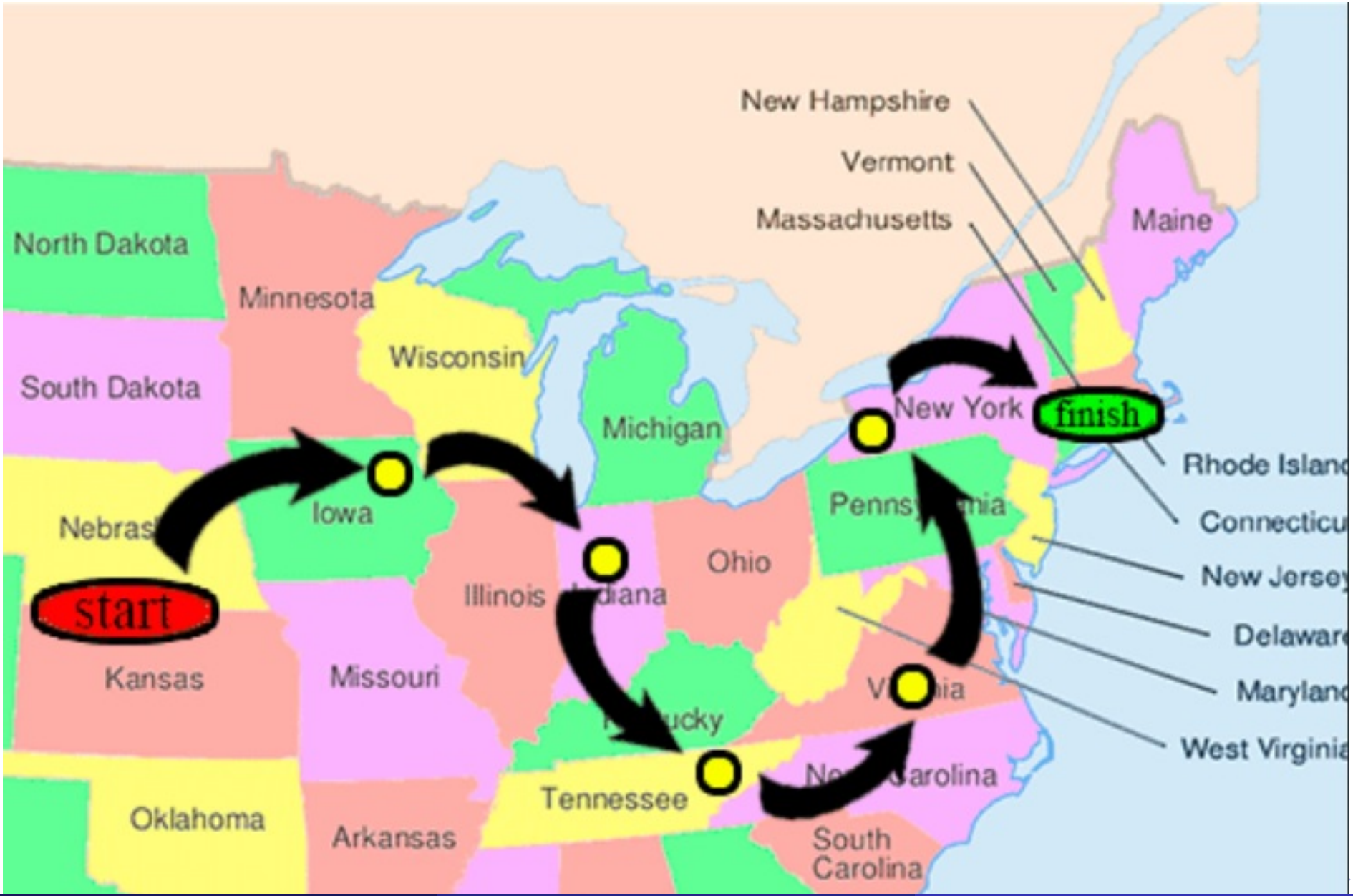
- Фрэнк Харари - стандартизация и расширение инструментария теории графов.
- 1959 г. - Эрдош и Реньи разрабатывают модель случайного графа.
- С 1960-х годов начинается активное развитие направления в общественных науках. Милгрэм проводит эксперимент по изучению реальных социальных сетей.
- С 1990-х годов - активное развитие направления физиками и информатиками.
- 1999 год - появление моделей малого мира (модель Ваттса-Строгатца) и модели предпочтительного присоединения (Барабаши-Альберт).
- После 2000 годов - развитие статистических моделей для анализа сетей (ERGM, SAOM), моделей распространениия информации в сетях и т.д. Особое внимание уделяется анализу больших сетей и больших данных.

Эксперимент 'Малый мир'

В 1967 году социолог Стэнли Милгрэм провел эксперимент с целью оценить число связей между двумя случайными людьми.

- Случайным людям из городов Омаха (Небраска) и Уичито (Канзас) были отправлены письма с описанием эксперимента и просьбой отправить письмо *целевому контакту* - человеку в Бостоне (Массачусетс);
- Если участники эксперимента лично знали целевой контакт, они могли отправлять ему письмо напрямую;
- Если участники эксперимента не знали целевой контакт, они должны были выбрать среди своих знакомых того, кто с наибольшей вероятностью был с ним знаком.

Эксперимент 'Малый мир'



Результаты эксперимента 'Малый мир'

- Из 296 стартовых писем финальной цели достигло 64 (29%);
- Цепочка от отправителя до получателя в среднем составила 5.5 человек. В дальнейшем это наблюдение переросло в заключение 'все в мире связаны между собой через шесть рукопожатий';
- Главным фактором для выбора 'посредников' стала географическая близость к целевому контакту.

В дальнейшем проводилось большое число схожих исследований на социальных онлайн-сетях (электронная почта, Facebook, MSN), в которых также было показано, что дистанция между двумя случайными вершинами социальной сети невелика и варьируется между 5 и 6.

Модели генерации сетей

Выработка моделей социальных сетей необходима для понимания механизмов формирования и функционирования сетей.

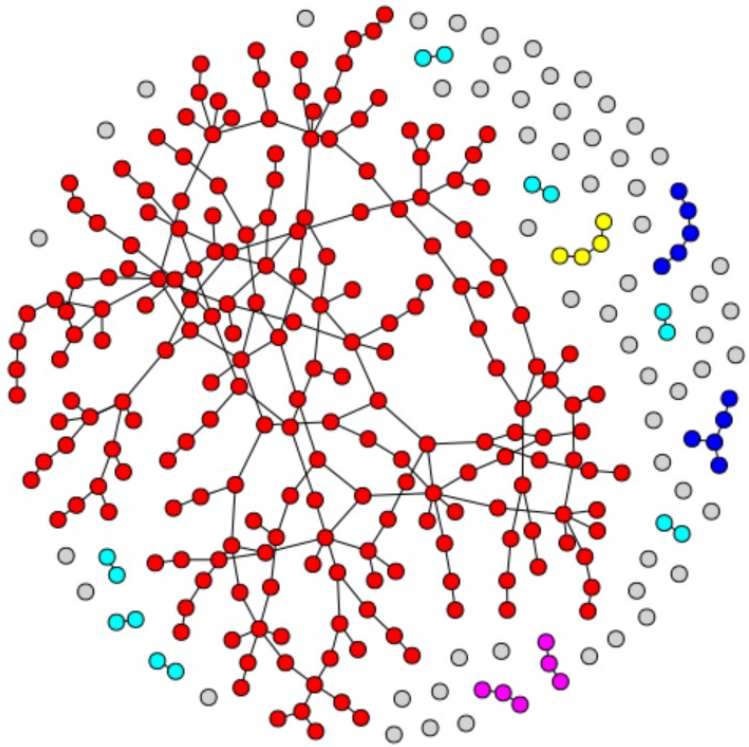
- Модель случайного графа Эрдоша-Реньи;
- Модель малого мира Ваттса-Строгатца;
- Модель предпочтительного присоединения Барабаши-Альберт.

Модель случайного графа

Модель случайного графа предложена Полом Эрдошем и Альфредом Реньи в 1959 году.

Алгоритм модели:

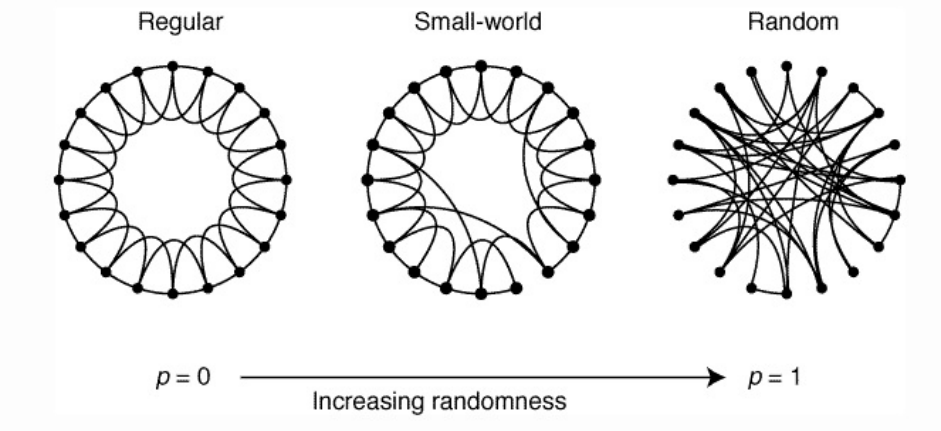
- Выбрано n изолированных вершин;
- Вершины последовательно соединяются ребрами случайным образом и независимо друг от друга.



Модель 'Малый мир'

В 1998 году Дунканом Ваттсом и Стивеном Строгатцем была предложена модель генерации сетей 'малый мир'.
Алгоритм модели:

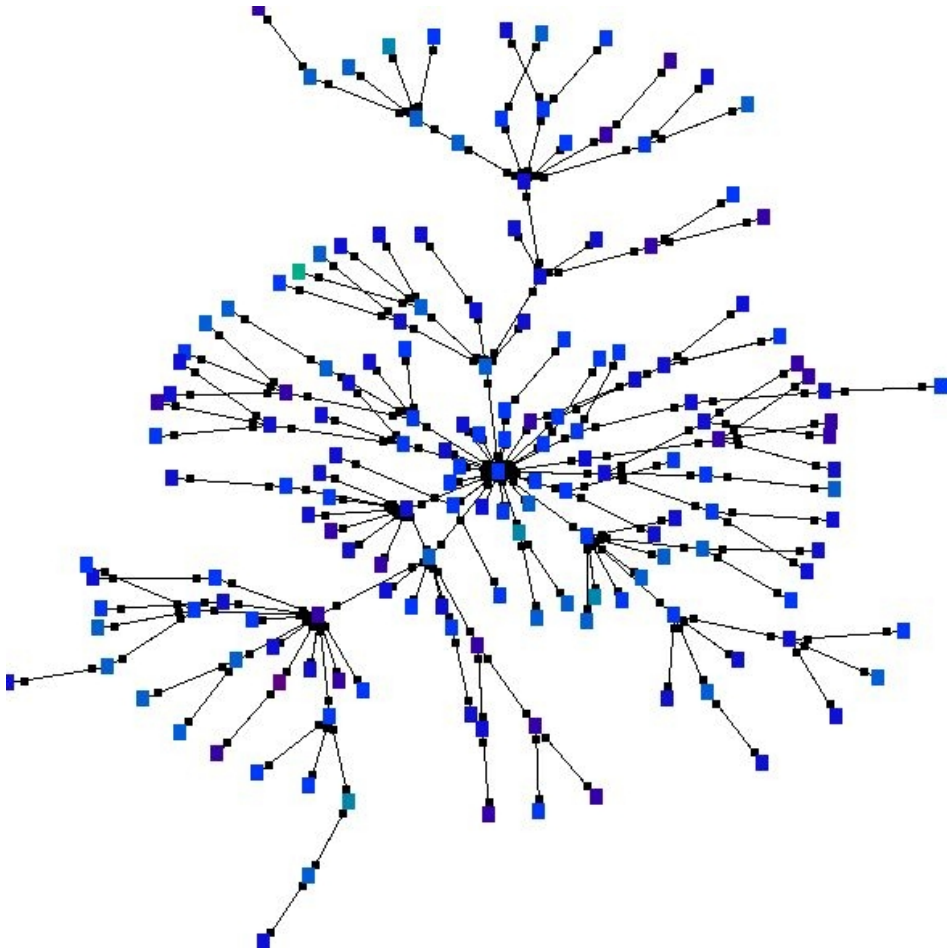
- Двумерная решетка, в которой вершины соединены только со своими ближайшими соседями;
- С определенной вероятностью p вершины разрывают связь с соседом и формируют ее со случайной вершиной.



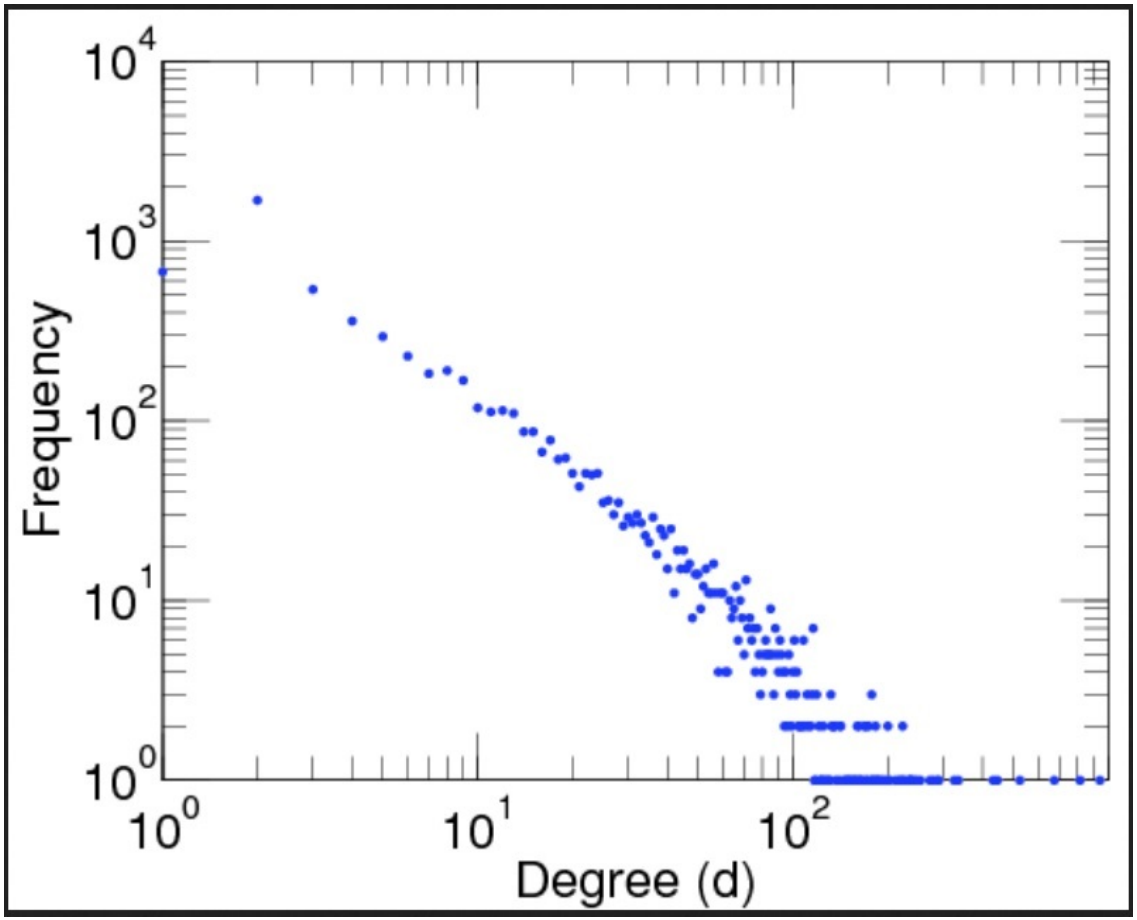
Watts, D. J., and Strogatz, S. H. (1998). Collective dynamics of 'small-world'networks. *Nature*, 393(6684), 440-442.

Степенной закон распределения центральностей

В эмпирических исследованиях показано, что некоторые вершины (хабы) имеют очень большое число связей, в то время, как основная масса вершин связана с очень небольшим числом соседей.



Степенной закон распределения центральностей



Модель предпочтительного присоединения

В 1999 году Альбертом-Ласло Барабаши и Рекой Альберт предложена модель генерации сетей по механизму предпочтительного присоединения. Основная идея модели в том, что чем больше связей у вершины, тем более предпочтительно для него формирование новых связей.
Алгоритм модели:

- Связная сеть, в которой все вершины имеют как минимум одну связь;
- В сеть по одному добавляются новые вершины, образуя связь с уже существующими. Вероятность присоединения к вершине прямо пропорциональна ее числу связей.

Barabasi, A. L., and Albert, R. (1999). Emergence of scaling in random networks. *Science*, 286(5439), 509-512.

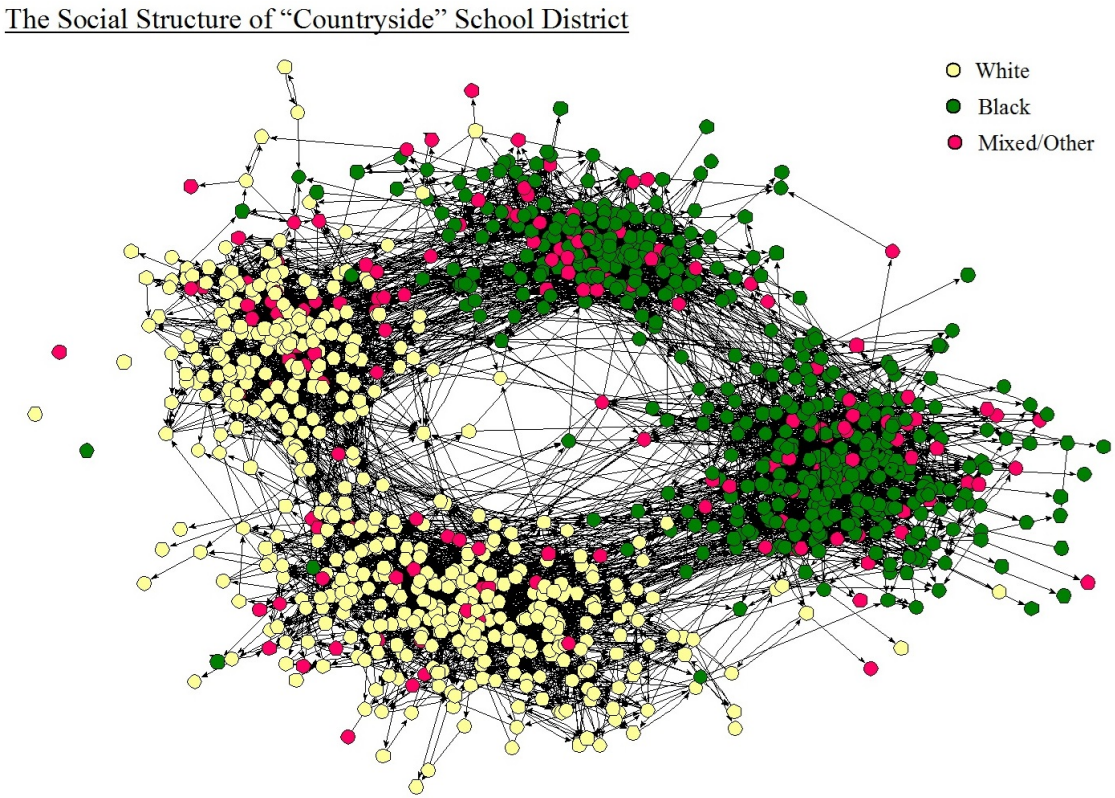
Структурные особенности социальных сетей

Таким образом, в совокупности для социальных сетей в той или иной степени характерны следующие отличительные особенности:

- Низкая плотность
- Высокая взаимность
- Высокая транзитивность
- Кластеризация
- Наличие небольшого количества **хабов**
- Гомофилия - склонность людей со схожими характеристиками быть связанными друг с другом

Сеть AddHealth

Сеть дружбы американских школьников. Номинировано до 5 друзей девочек и до 5 друзей мальчиков.



Высокие взаимность и транзитивность. Явно выявляются кластеры. Гомофилия по этнической принадлежности.

Динамика сетей

Для чего же нужен анализ структуры и динамики социальных сетей?
Выявление социальных и структурных механизмов, приводящих к формированию тех или иных социальных структур.
Какие могут быть механизмы: *взаимность, транзитивность, активность, популярность, гомофилия и т.д.*

Стохастические акторно-ориентированных модели

Предпосылки модели:

- Социальный индивидуализм - актор полностью осведомлен о позиции и характеристиках социальной системы, в которой он находится. Своими решениями о принятии/разрыве связей актор стремится оптимизировать свое положение в социальной сети.
- Акторы при принятии решений об изменениях социальных связей не координируют свои действия друг с другом.
- Изменения в социальной сети моделируются как **марковский процесс**, то есть положение актора в момент времени t зависит только от его положения в $t-1$. $t-2$, ... $t-n$ не оказывают влияния на структуру сети.
- Макро-изменения являются совокупностью микро-изменений в сети.

Snijders, T. A., Van de Bunt, G. G., and Steglich, C. E. (2010). Introduction to stochastic actor-based models for network dynamics. Social networks, 32(1), 44-60.

Целевая функция

Согласно модели каждый агент максимизирует свою целевую функцию.
$f_i(\beta, x) = \sum_k \beta_k s_{ki}(x)$
где $f_i(\beta, x)$ - целевая функция агента i в сети x, β_k - оценка параметра $s_{ki}(x)$. Оценки параметров интерпретируются как коэффициенты логистической регрессии.

Случаи, в которых САОМ используется

- Лонгитюдные данные по одной социальной группе
- Размер социальной сети до 1000 участников (сети дружбы и помощи в школе).
- Связи в социальных сетях - состояния, а не события. Например, связь - дружба, но не переписка отдельными сообщениями.

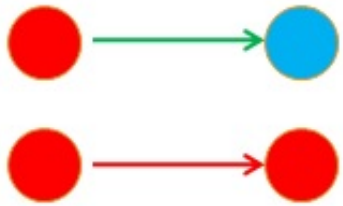
Таким образом, САОМ чаще всего используется для анализа взаимодействий в небольших организациях и в учебных учреждениях.

Коэволюция сетей и успеваемости студентов

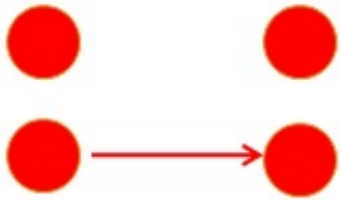
- Coleman et al. 1966 - социальное окружение оказывает важную роль в формировании индивидуальной результативности учащихся.
- Разные исследования по-разному оценивают процессы сообучения (Sacerdote 2000; Lyle 2007; Arcidiacono et al. 2012; Lin 2010). Результаты зависят от специфики выборки.
- Flashman (2012), Lomi et al. (2011) показали, что механизмы коэволюции сетей и достижений разнятся в зависимости от выборки.

Исследовательский вопрос

Каким образом происходит коэволюция социальных сетей и академической успеваемости учащихся.
Наблюдается ли эффект социального влияния, при котором акторы, связанные в социальной сети, с течением времени демонстрируют одинаковое поведение?



Наблюдается ли эффект социальной селекции, при котором акторы, схожие по успеваемости, формируют связи в социальных сетях?



Гипотезы исследования

Гипотеза 1: В социальной сети дружбы происходит социальная селекция - студенты склонны выбирать себе в качестве друзей сверстников со схожим уровнем академической успеваемости.
Гипотеза 2: В социальной сети дружбы происходит социальное влияние - студенты с течением времени перенимают успеваемость своих друзей.
Гипотеза 3: В социальной сети помощи происходит социальная селекция - студенты склонны выбирать себе в качестве помощников сверстников со схожим уровнем академической успеваемости.
Гипотеза 4: В социальной сети помощи происходит социальное влияние - студенты с течением времени перенимают успеваемость своих помощников.

- Гипотезы не противоречат друг другу, как социальная селекция, так и социальное влияние могут одновременно проходить (или не проходить) в системе.

Сбор данных

- Сбор опросных данных о характеристиках, сетях дружбы и помощи студентов. 3 волны в октябре 2013, феврале 2014 и июне 2014.
- *С кем Вы проводите больше всего свободного времени* - вопрос для определения сетей дружбы. Число номинацией неограничено.
- *К кому Вы обращаетесь за помощью в учебных вопросах* - вопрос для определения сетей дружбы. Число номинацией неограничено.
- *Галочкой отметьте тех, с кем Вы были знакомы до поступления* - вопрос для определения сетей дружбы. Число номинацией неограничено.
- Оценки получены из администрытивной базы данных университета. Система оценивания в университете открытая, результаты студентов по всем предматам и суммарный рейтинг доступны на сайте.

Описательная статистика сеть дружбы

Параметр	Первая волна	Вторая волна	Третья волна
Число студентов	117	117	117
Число связей	715	662	557
Плотность	0.053	0.049	0.041
Взаимность	0.63	0.60	0.51
Транзитивность	0.42	0.37	0.35
Коэфф. Жаккара	-	0.35	0.32

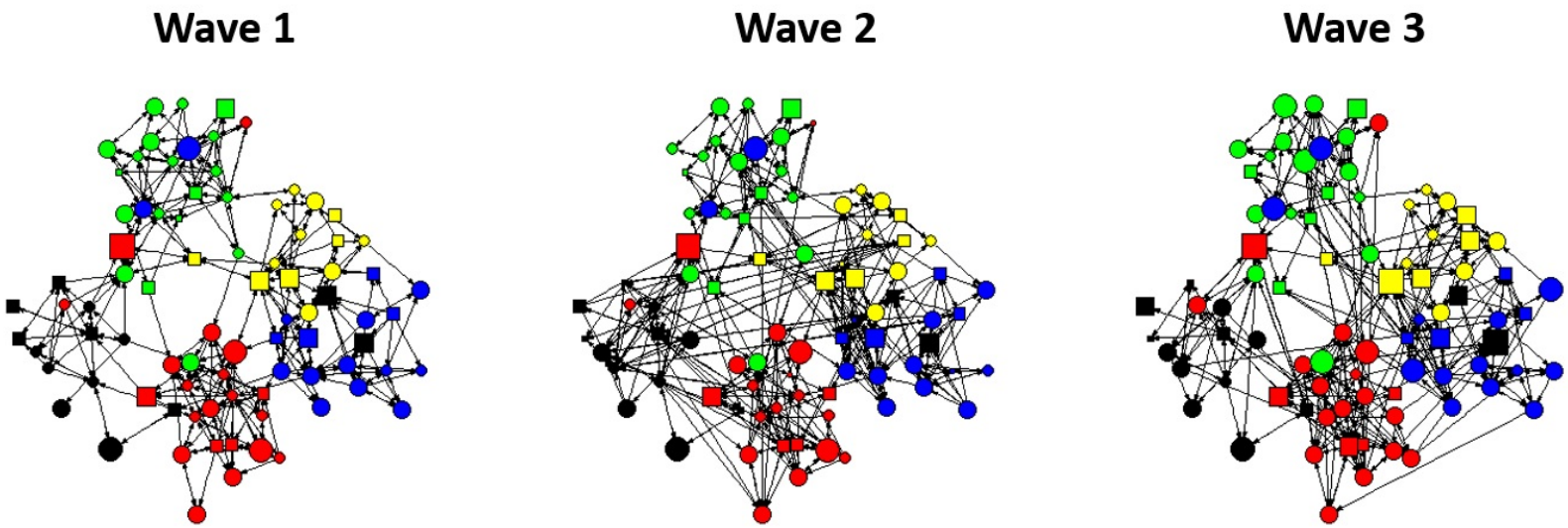
Описательная статистика сеть помощи

Параметр	Первая волна	Вторая волна	Третья волна
Число студентов	117	117	117
Число связей	226	267	248
Плотность	0.017	0.020	0.018
Взаимность	0.24	0.23	0.19
Транзитивность	0.29	0.28	0.27
Коэфф. Жаккара	-	0.28	0.26

Описательная статистика успеваемость

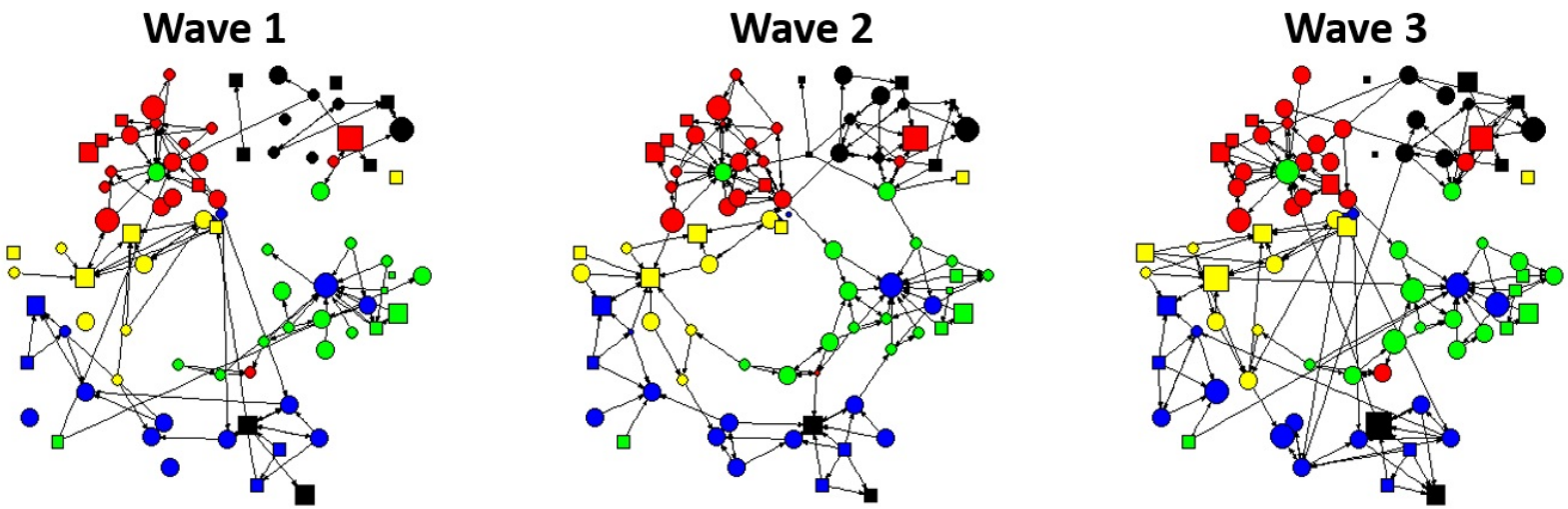
Параметр	Первая волна	Вторая волна	Третья волна
Среднее	2.47	2.39	2.79
Ст. отклонение	0.67	0.78	0.73
Maximum	4	4	4
Минимум	1	1	1
“Отличники”	7%	7%	14%
“Хорошисты”	35%	36%	55%
“Троечники”	55%	46%	26%
“Двоечники“	3%	11%	4%

Сеть дружбы



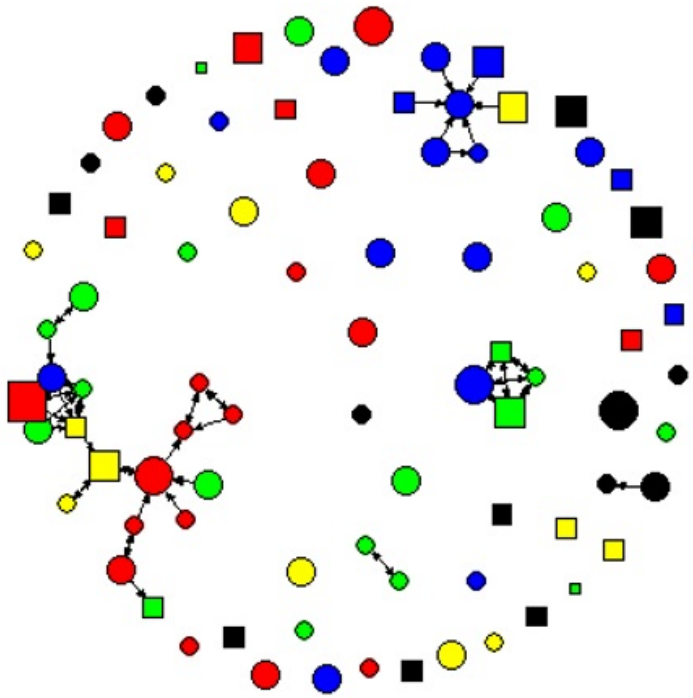
Вершины - студенты. Направленные связи между ними - сети дружбы. Студенты из одной учебной группы одного цвета. Девочки - кружочки. Мальчики - квадратики. Размер вершины пропорционален успеваемости.

Сеть помощи



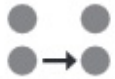
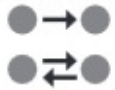
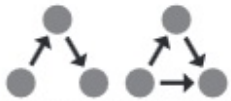
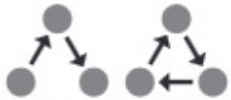
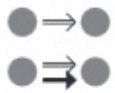
Вершины - студенты. Направленные связи между ними - сети помощи. Студенты из одной учебной группы одного цвета. Девочки - кружочки. Мальчики - квадратики. Размер вершины пропорционален успеваемости.

Сеть знакомства до зачисления

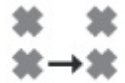


Вершины - студенты. Направленные связи между ними - сети знакомства. Студенты из одной учебной группы одного цвета. Девочки - кружочки. Мальчики - квадратики. Размер вершины пропорционален успеваемости.

Модель: оцениваемые эффекты 1

Название эффекта	Интерпретация	Иллюстрация
КОНТРОЛЬНЫЕ СЕТЕВЫЕ ЭФФЕКТЫ		
Плотность сети	Склонность акторов к формированию связей	
Взаимность	Склонность акторов к образованию взаимных связей	
ТРИАДНЫЕ ЭФФЕКТЫ		
Транзитивность	Склонность акторов устанавливать связи с теми, с кем устанавливают связи их друзья	
3-циклы	Склонность акторов к образованию циклических структур	
Посредничество	Склонность акторов занимать позиции посредников	
СВЯЗИ В ДРУГИХ СЕТЯХ		
Знакомство до поступления (связь в экзогенной сети)	Склонность акторов, знакомых до поступления в учебное заведение, завязывать связи дружбы	

Модель: оцениваемые эффекты 2

ЭФФЕКТЫ ПО АТРИБУТАМ (пол и успеваемость)		
Сходство по успеваемости/по полу	Склонность акторов со схожими характеристиками устанавливать связи между собой	
Экзогенные атрибуты (успеваемость) альтера	Склонность акторов с высокой успеваемостью быть более популярными	
Экзогенные атрибуты (успеваемость) эго	Склонность акторов с высокой успеваемостью быть более активными	
ЭФФЕКТЫ ДИНАМИКИ ПОВЕДЕНИЯ		
Линейный и квадратичный эффекты	Изменение показателя успеваемости в динамике	
Ассимиляция успеваемости	Склонность акторов перенимать успеваемость тех, с кем они связаны	
Популярность студентов с высоким средним баллом	Склонность популярных акторов демонстрировать более высокую успеваемость	
Активность студентов с высоким средним баллом	Склонность активных акторов демонстрировать более высокую успеваемость	

Результаты моделирования. Сеть дружбы 1

Параметр	Оценка (CO)	t-статистика
Параметр изменений 1	17.24*** (1.64)	0.03
Параметр изменений 2	16.08*** (1.43)	0.05
Контрольные эффекты		
Плотность	-2.02*** (0.14)	-0.03
Взаимность	1.69*** (0.11)	-0.05
Популярность	-0.03*** (0.01)	-0.02
Активность	-0.01 (0.01)	-0.01
Триадные эффекты		
Транзитивность	0.34*** (0.03)	-0.02
3-циклы	-0.33*** (0.06)	-0.03
Посредничество	-0.07* (0.03)	0.01
Связь в экзогенной сети		
Знакомство до поступления	0.95*** (0.14)	0.02
Обучение в одной группе	0.72*** (0.07)	-0.04
Связь в сети помощи	0.05*** (0.01)	-0.03

Результаты моделирования. Сеть дружбы 2

Эффект пола		
Пол альтера (1 – М)	0.13* (0.07)	0.02
Пол эго (1 – М)	0.21*** (0.07)	0.02
Схожесть по полу	0.24*** (0.06)	0.04
Эффекты успеваемости		
Успеваемость альтера	0.17*** (0.05)	0.02
Успеваемость эго	0.18*** (0.06)	0.03
Схожесть по успеваемости (селекция)	0.37 (0.21)	0.02
Динамика поведения		
Параметр изменений 1	0.54*** (0.13)	-0.00
Параметр изменений 2	1.14*** (0.20)	-0.02
Линейный эффект	1.01 (0.59)	0.01
Квадратичный эффект	0.38 (0.35)	-0.05
Ассимиляция успеваемости (влияние)	7.32*** (3.07)	0.03
Популярность	0.06 (0.11)	-0.01
Активность	-0.13 (0.13)	-0.01
Суммарная сходимость: 0.19		

Результаты моделирования. Сеть помощи 1

Параметр	Оценка (СО)	t-статистика
Параметр изменений 1	6.03*** (0.72)	0.02
Параметр изменений 2	5.84*** (0.64)	0.01
Контрольные эффекты		
Плотность	-3.24*** (0.29)	-0.01
Взаимность	1.06*** (0.19)	0.00
Популярность	0.06*** (0.02)	-0.04
Активность	-0.00 (0.03)	-0.01
Триадные эффекты		
Транзитивность	0.46*** (0.09)	0.00
3-циклы	-0.24 (0.20)	0.04
Посредничество	-0.29*** (0.08)	0.03
Связь в экзогенной сети		
Знакомство до поступления	1.01*** (0.21)	-0.02
Обучение в одной группе	1.40*** (0.13)	0.001
Связь в сети дружбы	0.13** (0.05)	0.02

Результаты моделирования. Сеть помощи 2

Эффект пола		
Пол альтера (1 – М)	0.07 (0.12)	-0.01
Пол эго (1 – М)	0.08 (0.13)	-0.00
Схожесть по полу	0.37*** (0.11)	0.01
Эффекты успеваемости		
Успеваемость альтера	0.77*** (0.26)	-0.01
Успеваемость эго	0.03 (0.25)	-0.01
Схожесть по успеваемости (селекция)	2.18*** (0.87)	-0.02
Динамика поведения		
Параметр изменений 1	0.63*** (0.15)	-0.01
Параметр изменений 2	1.48*** (0.32)	0.01
Линейный эффект	-0.56 (0.46)	0.06
Квадратичный эффект	-0.51*** (0.22)	-0.03
Ассимиляция успеваемости (влияние)	4.53 (2.74)	0.02
Популярность	0.38* (0.19)	-0.02
Активность	-0.15 (0.17)	0.03
Суммарная сходимость: 0.18		

*** p-value< 0.001, ** p-value<0.01, * p-value<0.05

Результаты 1

- В социальной сети дружбы наблюдается социальное влияние, то есть студенты склонны перенимать успеваемость своих друзей. Социальной селекции не зафиксировано.
- В социальной сети помощи наблюдается социальная селекция, то есть студенты склонны выбирать помощников со схожей результативностью. Социального влияния не зафиксировано.
- Отрицательная и значимая плотность - акторы формируют связи не случайным образом, а селективно.
- Сети взаимны и транзитивность важны для формирования связей.
- Совокупность положительной транзитивности и отрицательных 3-циклов для сети дружбы говорит о наличии локальной иерархии.
- Отрицательная популярность в сети дружбы - популярные акторы не становятся более популярными с течением времени.

Результаты 2

- Студенты не стремятся занимать позиции посредников в сетях.
- Студенты склонны формировать связи с теми, с кем они учатся в одной группе, были знакомы до поступления и связаны в сети помощи/дружбы.
- Мальчики более активны и популярны в сети дружбы.
- Студенты склонны дружить и просить о помощи студентов одного с ними пола.
- Студенты с высокими оценками более активны и популярны в сети дружбы.
- Студенты с высокими оценками более популярны в сети помощи.
- С течением времени студенты, с которым много обращаются за помощью повышают свои академические результаты.

Особенности селекции в сети помощи

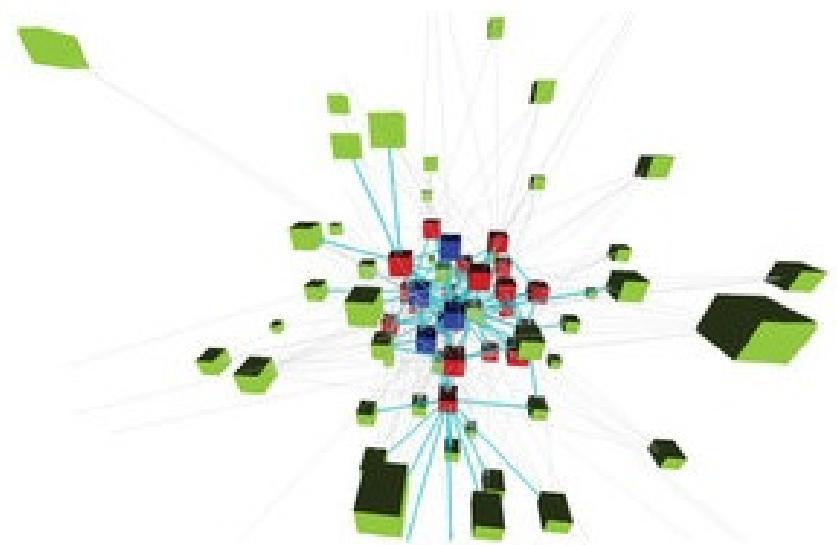
	Успеваемость альтера				
		Двоечник	Троечник	Хорошист	Отличник
Успеваемость эго	Двоечник	-0.63	-0.58	-0.54	-0.50
	Троечник	-1.32	0.17	0.22	0.26
	Хорошист	-2.02	-0.52	0.97	1.01
	Отличник	-2.72	-1.22	0.28	1.77

Особенности влияния в сети дружбы

	Успеваемость альтера			
	Двоечник	Троечник	Хорошист	Отличник
Успеваемость эго	Двоечник	1.17	-1.26	-3.70	-6.14
	Троечник	-0.99	1.45	-0.98	-3.42
	Хорошист	-2.40	0.04	2.48	0.04
	Отличник	-3.04	-0.60	1.84	4.28

Обсуждение результатов

Формирование Rich-club - сетевой структуры, состоящей из ядра и периферии постоянно удаляющихся друг от друга (Vaquero and Cebrian, 2013).



Спасибо за внимание!

Вопросы?