

## Student social network of friendship: structure and factors of formation

### Abstract

We consider the structure of social network of university students, and analyze factors that lead to the network formation. Using the data from student survey and administrative information, we estimate econometric models that assume homophily and geographic closeness effect on the likelihood of being friends. Specifically, we estimate linear probability models with and without fixed effects, and logistic model for probability of a tie between two students. The results confirm the significance of such friendship formation factors as sharing the same study group, living in the same dorm, closeness of academic achievement and gender homophily.

### Текст

В последнее время в научной литературе наблюдается повышенный интерес к вопросу социальных сетей в образовании, в том числе студенческих социальных сетей, что может объясняться совершенствованием методологии и инструментария анализа социальных сетей, возросшей доступностью эмпирических данных. Настоящая работа посвящена исследованию формирования социальных сетей в академической среде, а именно студенческих социальных сетей. Для студента дружеские связи являются не только неформальными отношениями в рамках института высшего образования, но и весьма полезным ресурсом, как во время, так и после окончания университета.

В литературе, посвященной изучению студенческих социальных сетей, выделяется несколько направлений. Во-первых, следует отметить работы, посвященные теоретическому моделированию сетей учащихся и оценке их характеристик, во-вторых, это работы, анализирующие влияния социальных сетей на академические успехи учащихся, так называемые эффекты среды, и, наконец, эмпирические работы по изучению формирования социальных сетей, оценке наиболее значимых факторов, определяющих возникновение связей в сети. Ниже будут рассмотрены работы, посвященные вопросу формирования социальных сетей в среде учащихся, в которых определяются ключевые факторы, оказывающих наибольшее воздействие на возникновение и структуру студенческих социальных сетей.

В существенном количестве работ, изучающих факторы формирования дружеских связей, в качестве одного из определяющих факторов указывается гомофилия, т.е. подобие черт. В случае студенческих сетей гомофилия проявляется в склонности учащихся дружить с близкими по различным характеристикам сверстниками.

Многие зарубежные исследователи обращаются к вопросу о влиянии этничности и расовой принадлежности на формирование дружеских связей среди студентов. Сравнивая группу американских и европейских студентов с группой студентов из Азии, Латинской Америки и африканских стран Кенни и Страйкер [Kenny, Stryker, 1994, 1996] пришли к выводу о том, что социальные сети формируются, в том числе, под влиянием расовой и этнической принадлежности. Данный эффект также известен как «расовая гомофилия». В результате данной работы, было обнаружено, что американские и европейские студенты создают более обширные и густые сети в отличие от своих однокурсников, принадлежащих другим этническим группам. Также было отмечено что, адаптация студентов в колледже происходит по-разному: студенты европеоидной расы получают большую поддержку от своих друзей, в то время как их сверстники из других этнических групп получают большую поддержку не от друзей, а от родственников и семьи.

Подобный эффект «расовой гомофилии» был также найден и в исследованиях Майера и Пуллера [Mayer, Puller, 2008] по данным социальной сети Facebook в США. В работе расовое единообразие определяется как детерминирующий фактор в процессе формирования дружеских связей учащихся. Более того, доказывается, что это характерно для разного типа учебных заведений, от частных закрытых школ, до крупных государственных университетов. Отмечается, впрочем, что политика университета, направленная на увеличение контактов между представителями разных этнических и расовых групп, может положительно повлиять на формирование дружеских социальных сетей среди этих студентов. В работе Вейнерта [Wejnert, 2010] также тестируется и находит подтверждение гипотеза о возможностях университетской среды влиять на формирование межнациональных и межрасовых связей среди студентов. Виммер и Льюис [Wimmer, Lewis, 2010] также обратили внимание на существенное влияние расового подобию на создание дружеской сети студентов. Однако авторы отмечают, существование таких факторов как транзитивность сети, и родственность студентов по другим характеристикам, в частности по социально-демографическому статусу.

Изучая природу гомофилии в студенческих дружеских сетях Коссинетс и Воттс [Kossinets, Watts, 2009] выяснили, что кроме схожести по социально-демографическим признакам, большое значение имеют студенческие триады, то есть транзитивность сети. Кроме того, авторы отмечают, что механизм выбора друзей в сети определяется не только схожестью характеристик, но и позицией потенциального друга, его влиятельностью в сети.

Транзитивность дружеских сетей учащихся подтверждается в исследовании социальных сетей школьников [Burk, et al., 2008]. Изучая социальные сети в динамике, авторы пришли к выводу, о существовании долгосрочных триад и когорт внутри сети учащихся, которые при переходе из начальной школы в среднюю школу могут временно разрушаться, но затем вновь восстанавливаются.

Моделируя формирование студенческих социальных сетей Мармарос и Сейсердот [Marmaros, Sacerdote, 2006] получили результаты, согласно которым ключевыми факторами, определяющими возникновение социальных сетей в студенческой среде являются принадлежность к одной этнической группе или расе, географическая близость, семейное окружение, а также общие интересы в рамках обучения в университете, такие как изучаемые курсы или участие в университетских спортивных секциях. Один из основных результатов данного исследования свидетельствует о том, что вероятность возникновения дружбы между студентами одной расовой группы значительно выше, чем между студентами из разных расовых групп. Однако, авторы полагают, что совместное обучение со сверстниками из разных этнических групп, особенно на первом году обучения, увеличивает шансы на формирование интернациональных дружеских связей.

В российских вузах вопросы расы и этничности не столь злободневны как в США. Тем не менее, после введения ЕГЭ в ведущих вузах возросло число студентов из других регионов, ежегодно увеличивается поток иностранных студентов, в том числе из стран СНГ и Балтии [Гаврилова, Яценко, 2012].

Наблюдается также существование гомофилии по другим атрибутивным характеристикам, таким как социально-экономический статус учащихся, схожесть хобби и интересов, выбор одинаковых учебных траекторий. В исследовании по данным немецкой социальной сети StudyVZ отмечается, что студенты технических и математических направлений с большей вероятностью формируют дружеские связи со студентами схожих направлений, чем со студентами гуманитарных и художественных университетов [Lee, Scherngell, Barber, 2011]. В другом исследовании найдено, что для студентов первого курса детерминирующей является схожесть по половому и расовому признакам, в то

время как для более старших курсов возрастает значимость общих интересов, например религиозные убеждения, членство в определенных клубах [Godley, 2008].

В работе Элиас, Элиас и Ронкони [Elias, Elias, Ronconi, 2007] анализирующей данные по учащимся Аргентины, делаются выводы о значимости таких факторов как успеваемость студентов и внешняя привлекательность. Согласно исследованию у внешне привлекательного учащегося с высокими академическими достижениями вероятность сформировать дружеские связи выше, чем у тех, кто не обладает подобными качествами. Авторы также тестируют модель для школ с совместным обучением, так и для школ с раздельным обучением. В результате делается заключение о значимости академических успехов для обоих типов школ, в то время как внешняя привлекательность оказывается важным фактором только для смешанных школ.

Исследование Соэтевента и Куремана [Soetevent, Kooreman, 2005] также приводит доказательства влияния факторов сходства на вероятность возникновения дружеских связей. Значимость отдельных параметров варьируется в зависимости от типов поведения студентов. Так, например, в случае личных встреч большее влияние оказывают факторы этнического и расового характера, тогда как в случае он-лайн коммуникации или телефонных звонков более значимым оказывается фактор гендерной принадлежности.

Кроме определения механизма формирования дружеских связей среди студентов за счет детерминантов, связанных с единообразием и подобием характеристик учащихся, в литературе находятся убедительные свидетельства влияния факторов географической близости студентов, называемые «эффект близости» или «эффект соседства». Эффект близости, или соседства, связан с совместным проживанием студентов в одной комнате в общежитии и географической близостью между местами их проживания и обучения [Festinger, et al., 1950].

Анализируя данные германской социальной сети StudyVZ, исследователи пришли к выводу о непосредственном влиянии географического местоположения на формирование дружеских связей среди студентов разных вузов Германии [Lee, Scherngell, Barber, 2011]. Согласно полученным результатам удаленность, соответствующая 100 минутам езды на автомобиле уменьшает вероятность формирования дружеских связей на 91%, также дружеские связи чаще формируются среди студентов, чьи университеты находятся на одной федеральной области.

К аналогичным выводам относительно значимости географического расстояния в процессе формирования дружеских связей пришли Пресиадо с соавторами [Preciado, et al., 2012]. Исследователи однозначно определили зависимость между вероятностью возникновения дружбы и географической удаленностью двух учащихся. Кроме того, для учеников из разных учебных заведений такая зависимость более сильная, чем для тех, кто обучается в одной школе. Авторы также показали, что данный эффект географической близости не связан с транзитивностью студенческой сети.

В другой работе [Traud, et al., 2011], анализирующей данные социальной сети Facebook, также подтвердилось предположение о воздействии эффекта близости на формирование дружеских связей учащихся. Однако авторы работы утверждают, что не существует одного детерминирующего фактора, определяющего механизм формирования дружеских связей и только совместное влияние нескольких характеристик (совместное проживание, обучение на одном курсе или выбор одинаковых предметов) может объяснить структуру студенческой социальной сети.

В исследовании Фостер [Foster, 2005] также делается вывод о том, что процесс формирования дружеских связей среди студентов зависит от совокупности факторов, относящихся как к эффектам гомофилии, таких как раса, пол, успеваемость, так и к эффектам географической близости (например, обучение в одном кампусе).

В работе Валеевой, Польдина и Юдкевич [Валеева, Польдин, Юдкевич, 2013] для оценки вероятности образования связей между студентами одного из факультетов российского вуза использовался метод экспоненциального моделирования случайных графов (Exponential random graph model). Связи дружбы образуются чаще между студентами, обучающимися в одной группе и принадлежащими к одному и тому же полу. Чем меньше разница в среднем балле между двумя студентами, тем больше вероятность образования между ними связей.

Таким образом, анализируя социальную сеть в качестве зависимой переменной, исследователи выделяют два основных механизма её формирования: гомофилия и «эффект соседства». В литературе по данной тематике прослеживаются убедительные свидетельства того, что студенты предпочитают дружить со сверстниками из собственной этнической и расовой группы, со схожими социально-экономическими характеристиками и интересами. Выбор одинаковых программ обучения, равно как и участие в одних и тех же университетских обществах увеличивает вероятность возникновения дружеских связей между студентами. Кроме перечисленных выше факторов значение имеет и географическая удаленность: чем ближе студенты живут друг к другу, тем вероятней они будут находиться в одной социальной сети.

В данном исследовании анализируются основные характеристики студенческих социальных сетей и оцениваются факторы, определяющие их формирование, на примере студентов первого курса одного российского университета. Также рассматривается структура дружеских социальных сетей учащихся. Выборка студентов была сформирована в результате опроса студентов трех факультетов.

#### *1. Данные и основные характеристики сети*

В рамках данного исследования был проведен опрос студентов трех факультетов первого курса НИУ ВШЭ – Нижний Новгород: бизнес-информатики, менеджмента и экономики. Анкета, предложенная студентам, позволила получить информацию о социально-экономическом статусе респондентов, их академических характеристиках и о предпочтениях в проведении досуга. Для построения сетей студентам были заданы два вопроса, в которых респонденты должны были указать фамилии своих однокурсников, с которыми они чаще всего общаются.

Для описания сети в данном разделе используется ненаправленные связи: два студента считаются друзьями, если хотя бы один из них указал в анкете другого студента в качестве друга. Графическое представление дружеской сети первокурсников представлено в приложении.

При исследовании структуры сетей дружбы студентов использовались стандартные сетевые характеристики. Результаты анализа представлены в таблице 1.

Таблица 1. Основные сетевые характеристики дружеских сетей первокурсников

Показатели	Факультет экономики	Факультет менеджмента	Факультет бизнес информатики
Число узлов	105	97	93
Число связей	356	333	309
Диаметр	8	7	10
Плотность (%)	6,5	7,2	7,2
Средняя степень	6,8	6,9	6,6
Среднее расстояние между достижимыми парами узлов	3,7	3,2	4,3
Коэффициент кластеризации сети (%)	52,5	45,2	43,3
Промежуточная центральность	145,5	106,1	153,9

Показатель «число узлов» в данном случае – это число студентов на каждом из анализируемых факультетов, входящих в сеть дружбы. Практически все студенты, на момент опроса обучавшиеся на первом курсе, попали в сеть дружбы своих однокурсников. Только девять студентов не попали в сети дружбы (один с факультета экономики, три с факультета менеджмента и пять с факультета бизнес информатики), данные студенты никого не отметили в качестве своих друзей и не были отмечены другими учащимися в качестве друзей.

Число связей показывает количество дружеских связей на факультетах. Для описания структуры сетей используется показатель плотности, который рассчитывается на основании данных о числе связей и узлов в сети. Плотность сети (density) определяется как отношение связей в сети к количеству всех возможных связей для заданной сети [Wasserman, Faust, 1994]:

$$Density = \frac{L}{N(N-1)/2}, \quad (1)$$

где  $L$  – наблюдаемое количество связей,  $N$  – число узлов в сети. Для анализируемых сетей данный показатель составляет около 7% (стандартное отклонение 0,003), то есть реализуется менее 10% всех возможных взаимодействий.

Диаметр и среднее расстояние между достижимыми парами узлов в анализируемых сетях также варьируются незначительно. Данные характеристики дают общее представление об удаленности студентов друг от друга и скорости распространения информации в дружеской сети. Диаметр (diameter) показывает, насколько узлы сети близки друг к другу. Формально диаметр – это самая длинная протяженность, то есть максимальное число уровней связей разделяющих двух любых студентов. Показатель расстояние (distance) показывает наиболее короткий путь между любыми двумя точками в сети, то есть минимальное число уровней связей разделяющих двух любых студентов. Среднее расстояние между двумя достижимыми парами узлов можно рассчитать по формуле [Sanjeev, 2007]:

$$Distance = \frac{\sum_{i \in N} \sum_{j \in N} d(i, j; g)}{n(n-1)}, \quad (2)$$

где  $d(i, j; g)$  – расстояния между двумя узлами  $i$  и  $j$  в сети  $g$ .

Таким образом, на факультете менеджмента любым двум студентам нужна в среднем цепочка из 3 студентов, чтобы связаться друг с другом, максимум – 7 учащихся, на факультетах экономики и бизнес информатики средняя длина такой цепочки – 4 человека, максимальная – 8 и 10 учащихся, соответственно.

Степень (degree) представляет собой количество связей, которым обладает узел в сети [Wasserman, Faust, 1994], то есть это сумма всех соседних узлов связанных с заданным. Для узла  $i$ :  $d_i(g) = |N_i(g)|$ . Степень может быть рассчитана для каждого узла внутри сети, либо для всей сети, в таком случае говорят о средней степени. Средняя степень в сети рассчитывается по формуле:

$$\text{Average Degree} = \sum_{i \in N} \frac{d_i(g)}{n}, \quad (3)$$

где  $d_i(g)$  – значение степени для каждого узла.

В случае дружеской сети степень представляет собой число друзей одного человека. Средняя степень характеризует среднее число друзей, приходящихся на одного человека в сети. Согласно результатам, студенты трех анализируемых факультетов имеют в среднем по 7 друзей из числа своих однокурсников.

Показатель степени, является важной характеристикой, отражающей статус студента в дружеской сети. Можно предположить, что чем больше друзей есть у студента, тем он популярней и влиятельней в своей социальной сети. На рисунке 1 представлено распределение участников трех дружеских сетей по степени, т.е. по числу дружеских связей.

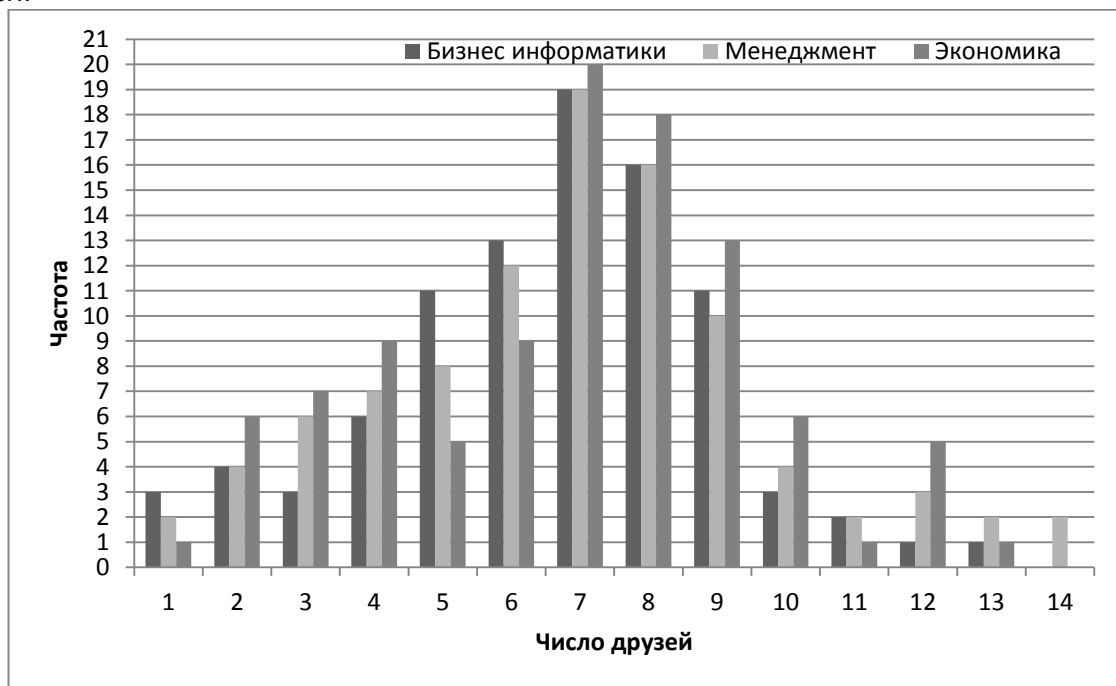


Рис. 1. Распределение степени в сетях дружбы по трем факультетам

Как видно из графика дружеские сети практически идентичны по характеру распределения числа друзей среди участников сети. Для трех студентов факультетов характерна дружба с 6-9 однокурсниками. На самое большое число друзей (14 однокурсников) отмечается у студента факультета менеджмента, по 13 друзей есть у одного учащегося с факультета бизнес информатики, у одного учащегося с факультета экономики и двух учащихся с факультета менеджмента.

Коэффициент кластеризации показывает тенденцию актов социальной сети формировать более тесные подгруппы, то есть кластеры. Склонность людей объединяться в группы в рамках общей социальной сети может частично объясняться эффектом гомофилии, подробно описанном в литературе, согласно которому контакты между похожими людьми возникают с большей вероятностью, чем между людьми с разными характеристиками. Формально коэффициент кластеризации, рассчитанный для узла сети, показывает, насколько близки соседние связанные узлы к формированию клики или полного графа. В отношении дружеских сетей студентов коэффициент кластеризации сети может быть интерпретирован как стремление учащихся объединяться в более узкие, закрытые группы. Коэффициент кластеризации для каждого узла в сети можно рассчитать по формуле [Hardiman, Katzir, 2013]:

$$Clustering\ Coefficient = \frac{2l_i}{d_i(d_i - 1)}, \quad (4)$$

где  $l_i$  – число триплетов (трех связанных узлов) для данного узла,  $d_i$  – степень данного узла. Другими словами, коэффициент кластеризации показывает отношение числа реально существующих триплетов к числу потенциально возможных триплетов. Расчет коэффициента кластеризации сети есть среднее арифметическое по всем индивидуальным коэффициентам кластеризации. Для анализируемых сетей коэффициент кластеризации варьируется от 43,3 до 52,5%, то есть около половины учащихся стремятся к формированию кластеров в рамках дружеских связей на факультете.

При описании структуры социальной сети большое внимание уделяется анализу положения узлов в сети. Нахождение «центральных» узлов сети представляется одной из основных задач анализа социальных сетей [Freeman, 1978; Vonacich, 1987; Opsahl, et al., 2010]. Одним из наиболее эффективных инструментов определения позиции узла в сети признается центральность (centrality) [Ajith, et al., 2010]. Центральность характеризует позицию узла по отношению к остальным узлам в данной сети [Friedkin, 1991]. При оценке промежуточной центральности анализируется положение узла не сточки зрения количества связей, которые он имеет, а с точки зрения числа наиболее коротких путей, проходящих через данный узел. Таким образом, промежуточная центральность показывает положение узла между другими узлами, то есть узел выступает неким связующим звеном. Узел сети, через который проходит наибольшее количество самых коротких путей, имеет наибольшую промежуточную центральность.

Для расчета показателя промежуточной центральности узла сначала считаются все самые короткие пути между парами узлов в сети. Затем для каждой пары узлов определяется количество кратчайших путей, которые проходят через узел, для которого определяется значение промежуточной центральности. Наконец, находится сумма отношения числа кратчайших путей проходящих через анализируемый узел ко всем кратчайшим путям в сети. Формально, значение показателя промежуточной центральности может быть рассчитано по формуле, предложенной Фриманом [Freeman, 1978]:

$$C_b(v) = \sum_{s \neq t \neq v \in V} \frac{\sigma(s, t | v)}{\sigma(s, t)}, \quad (5)$$

где  $\sigma(s, t)$  – сумма всех кратчайших путей между вершинами  $s$  и  $t$ ,  $\sigma(s, t | v)$  – сумма кратчайших путей между вершинами  $s$  и  $t$ , проходящих через узел  $v$ . В качестве промежуточной центральности всей сети используется среднее арифметическое значение индивидуальных показателей промежуточной центральности. Так как число вершин и связей в трех анализируемых сетях незначительно варьируется, показатель

промежуточной центральности можно не нормировать. Сравнение сетей по среднему показателю центральности нецелесообразно, однако с помощью показателя промежуточной центральности можно выявить наиболее значимых студентов в дружеских сетях. Анализ индивидуальной промежуточной центральности показал, что для большинства узлов значение промежуточной центральности варьируется от 0 до 100. Медианные значения равны 45,6, 78,3, 84,1 соответственно для факультетов бизнес информатики, менеджмента и экономики. Однако в сетях есть узлы с показателем промежуточной центральности значительно выше медианного и среднего уровня, можно предположить, что такие учащиеся играют важную роль в своих сетях дружбы с точки зрения распространения информации.

Анализ сетевых характеристик показал, что все анализируемые сети имеют схожую структуру. В среднем у студентов есть 6-7 друзей из числа однокурсников, для связи с любым студентом внутри сети нужно в среднем от 3 до 4 уровней общих знакомых, максимум 10 – для дружеской сети факультета бизнес информатики. Около половины всех учащихся формируют кластеры внутри своих дружеских сетей.

При оценке факторов, влияющих на вероятность формирования дружеских связей в социальных сетях студентов, были использованы предположения о существовании эффекта гомотилии и эффекта соседства. Описательная статистика данных представлена в таблице 2.

Таблица 2. Описательная статистика выборки.

	Число наблюдений	Среднее	Стандартное отклонение	Минимум	Максимум
Факультет бизнес информатики					
Средний балл	96	6,80	1,09	5	9,31
% мужчин	98	56,1			
% проживающих в общежитии	83	9,6			
Оценка благосостояния	72	2,88	0,73	1	4
Факультет менеджмента					
Средний балл	100	7,08	0,87	4,65	9,15
% мужчин	100	29			
% проживающих в общежитии	87	21,8			
Оценка благосостояния	75	2,76	0,59	1	4
Факультет экономики					
Средний балл	104	7,12	1,04	5,25	9,22
% мужчин	106	28,3			
% проживающих в общежитии	100	16			
Оценка благосостояния	83	3,02	0,60	1	4

Средний балл вычислялся на основании рейтингов учащихся, публикуемых деканатами соответствующих факультетов. Публикация рейтингов происходила после проведения опроса, т.е. респонденты на момент опроса не знали о величине среднего балла своих однокурсников. Значение среднего балла по всем факультетам варьируется незначительно в пределах от 6,8 до 7,12, что соответствует оценке «хорошо». В данном случае учитывались данные всех учащихся, в том числе тех, кто не участвовал в опросе. Переменная «Оценка благосостояния» является дискретной и принимает значение от 1 – малообеспеченные до 4 – хорошо обеспеченные. В данном случае респонденты сами определяли свое материальное положение. Большинство респондентов, ответивших на



вопрос о своем материальном благосостоянии, оценивают его как «в целом, обеспеченные». Следует отметить, что многие участники опроса предпочли не отвечать на этот вопрос. Гендерное распределение студентов определялось с использованием данных, предоставляемых деканатами университета. На факультете экономики и факультете менеджмента среди студентов большую долю составляют девушки. Доля мужчин в среднем варьируется около 30%. На факультете бизнес-информатики доля мужчин чуть выше, чем доля девушек, 56,1%. Процент учащихся, проживающих в общежитии, он определялся на основании ответов респондентов и рассчитывался как доля учащихся, проживающих в общежитии, ко всем учащимся, отметившим свое место проживания (с родителями, родственниками и пр.). Таким образом, вероятно реальная доля студентов, проживающих в общежитии, может отличаться от доли, полученной при анализе, так как некоторые респонденты проигнорировали данный вопрос.

## 2. Методология и результаты оценивания

Мы предполагаем, что каждый студент может формировать дружеские связи с любым другим однокурсником своего факультета. Каждое наблюдение относится к паре, составленной из двух студентов одного факультета. Число наблюдений, таким образом, соответствует числу всевозможных пар.

Зависимая бинарная переменная описывает статус дружеской связи:  $F_{ij}=1$ , если студенты  $i$  и  $j$  являются друзьями,  $F_{ij} = 0$  – если нет. Для анализа мы использовали ненаправленные связи. Вероятность дружбы предполагается зависящей от индивидуальных характеристик каждого студента и характеристик пары. Зависимость от индивидуальных характеристик будем учитывать с помощью фиксированных эффектов,  $\vartheta_i$  и  $\vartheta_j$ . Это позволяет учесть не только наблюдаемые, но и ненаблюдаемые характеристики студентов и избежать смещения оценок вследствие пропущенных переменных.

Что касается зависимости от характеристик пары, мы опираемся на гипотезу гомофилии в образовании связей и предполагаем дружбу функцией от близости наблюдаемых индивидуальных характеристик. Для описания близости таких дискретных индивидуальных характеристик, как пол, принадлежность к конкретной студенческой группе и оценка материального благосостояния, мы вводим бинарные переменные-индикаторы, принимающие значение 1 в случае, совпадения этих характеристик у рассматриваемой пары студентов, и значение 0 в случае, если эти характеристики различаются:

$$D(x^{ik}, x^{jk}) = I(x^{ik} = x^{jk}). \quad (6)$$

Для среднего балла степень близости описывается через абсолютные значения разности индивидуальных значений:

$$D(x^{ik}, x^{jk}) = -|x^{ik} - x^{jk}|. \quad (7)$$

С учетом введенных обозначений, оцениваемая эконометрическая модель имеет вид:

$$p_t = Prob(F_{ij} = 1) = \beta_0 + \sum_{k=1}^K \beta_k D(x_t^{ik}, x_t^{jk}) + \theta^i + \theta^j = \mathbf{D}'_t \boldsymbol{\beta} + \theta^i + \theta^j, \quad t = 1, \dots, N. \quad (8)$$

Из-за того, что оценивается зависимость с двумя фиксированными эффектами, мы используем линейную вероятностную модель, а не нелинейные модели типа логистической. Фиксированные эффекты в эконометрической модели учитываются с помощью переменных индикаторов для каждого студента в паре (т.н. метод наименьших квадратов с фиксированными переменными (Магнус, Катывшев, Пересецкий 2007)).

Оценки моделей вероятности возникновения дружбы между студентами представлены в таблицах 3-5 для факультетов бизнес-информатики, менеджмента и экономики, соответственно. В столбцах 1 и 2 таблиц приведены оценки модели, заданной

уравнением (8). В столбцах 3 и 4 приведены оценки модели без фиксированных эффектов.

Для сравнения мы также оценили логистическую модель

$$p_i = 1 / (1 + e^{-D_i \beta}). \quad (9)$$

В силу малой информативности значений коэффициентов, в столбцах 5 и 6 показаны значения средних предельных эффектов от изменения соответствующей зависимой переменной. Для непрерывной переменной (близость успеваемости) средний предельный эффект рассчитывается как среднее выборочное

$$AME_k = \frac{1}{N} \sum_{i=1}^N \frac{\partial p_i(D_i \hat{\beta})}{\partial D_i} = \frac{1}{N} \sum_{i=1}^N \frac{e^{D_i \hat{\beta}}}{(1 + e^{D_i \hat{\beta}})^2} \hat{\beta}_k. \quad (10)$$

При расчете среднего предельного эффекта для дискретных переменных производная в (5) заменяется разностью:

$$AME_k = \frac{1}{N} \sum_{i=1}^N (p_i(D_i \hat{\beta} | D_k = 1) - p_i(D_i \hat{\beta} | D_k = 0)). \quad (11)$$

Таблица 3. Оценки линейной и логистической моделей вероятности возникновения дружбы, факультет бизнес информатики.

Независимые переменные	Линейная модель			Логистическая модель, предельные эффекты		
	(1)	(2)	(3)	(4)	(5)	(6)
Обучение в одной группе	0,268*** (0,014)	0,280*** (0,012)	0,264*** (0,014)	0,275*** (0,012)	0,167*** (0,008)	0,179*** (0,008)
Одинаковый пол	0,025*** (0,006)	0,029*** (0,005)	0,025*** (0,006)	0,030*** (0,005)	0,027*** (0,006)	0,031*** (0,005)
Проживание в одном общежитии	0,034** (0,016)	0,029* (0,015)	0,030*** (0,007)	0,026*** (0,007)	0,031*** (0,008)	0,028*** (0,007)
Близость среднего балла	0,023*** (0,005)	0,023*** (0,004)	0,018*** (0,003)	0,019*** (0,003)	0,021*** (0,004)	0,022*** (0,003)
Близость материального положения	0,012 (0,009)		0,020*** (0,006)		0,019*** (0,006)	
Константа	-0,026 (0,034)	-0,019 (0,033)	-0,018** (0,008)	-0,007 (0,007)		
Фиксированные эффекты	да	да	нет	нет	нет	нет
Число наблюдений	4970	6480	4970	6480	4970	6480
R <sup>2</sup>	0,238	0,244	0,208	0,219		
Псевдо R <sup>2</sup>					0,363	0,376

В скобках указаны стандартные погрешности

\*  $p < 0,1$ , \*\*  $p < 0,05$ , \*\*\*  $p < 0,01$

Таблица 4. Оценки линейной и логистической моделей вероятности возникновения дружбы, факультет менеджмента.

Независимые переменные	Линейная модель			Логистическая модель, предельные эффекты		
	(1)	(2)	(3)	(4)	(5)	(6)
Обучение в одной группе	0,217 <sup>***</sup> (0,011)	0,196 <sup>***</sup> (0,009)	0,217 <sup>***</sup> (0,011)	0,198 <sup>***</sup> (0,010)	0,164 <sup>***</sup> (0,009)	0,147 <sup>***</sup> (0,007)
Одинаковый пол	0,066 <sup>***</sup> (0,009)	0,060 <sup>***</sup> (0,007)	0,036 <sup>***</sup> (0,006)	0,032 <sup>***</sup> (0,005)	0,039 <sup>***</sup> (0,007)	0,034 <sup>***</sup> (0,006)
Проживание в одном общежитии	0,061 <sup>***</sup> (0,011)	0,054 <sup>***</sup> (0,010)	0,029 <sup>***</sup> (0,006)	0,030 <sup>***</sup> (0,005)	0,027 <sup>***</sup> (0,007)	0,028 <sup>***</sup> (0,006)
Близость среднего балла	0,032 <sup>***</sup> (0,007)	0,035 <sup>***</sup> (0,005)	0,017 <sup>***</sup> (0,004)	0,019 <sup>***</sup> (0,003)	0,018 <sup>***</sup> (0,004)	0,021 <sup>***</sup> (0,004)
Близость материального положения	-0,006 (0,008)		0,001 (0,006)		0,003 (0,006)	
Константа	0,026 (0,042)	0,018 (0,036)	-0,014 <sup>**</sup> (0,007)	-0,007 (0,005)		
Фиксированные эффекты	да	да	нет	нет	нет	нет
Число наблюдений	5550	7482	5550	7482	5550	7482
$R^2$	0,196	0,174	0,158	0,140		
Псевдо $R^2$					0,285	0,263

В скобках указаны стандартные погрешности

\*  $p < 0,1$ , \*\*  $p < 0,05$ , \*\*\*  $p < 0,01$

Таблица 5. Оценки линейной и логистической моделей вероятности возникновения дружбы, факультет экономики.

Независимые переменные	Линейная модель			Логистическая модель, предельные эффекты		
	(1)	(2)	(3)	(4)	(5)	(6)
Обучение в одной группе	0,176 <sup>***</sup> (0,011)	0,154 <sup>***</sup> (0,008)	0,169 <sup>***</sup> (0,011)	0,149 <sup>***</sup> (0,008)	0,115 <sup>***</sup> (0,006)	0,101 <sup>***</sup> (0,005)
Одинаковый пол	0,023 <sup>***</sup> (0,008)	0,036 <sup>***</sup> (0,006)	0,034 <sup>***</sup> (0,005)	0,038 <sup>***</sup> (0,004)	0,038 <sup>***</sup> (0,007)	0,041 <sup>***</sup> (0,005)
Проживание в одном общежитии	0,022 <sup>*</sup> (0,012)	0,016 <sup>*</sup> (0,009)	0,010 (0,006)	0,008 (0,005)	0,010 <sup>*</sup> (0,006)	0,007 (0,005)
Близость среднего балла	0,034 <sup>***</sup> (0,004)	0,031 <sup>***</sup> (0,003)	0,028 <sup>***</sup> (0,003)	0,027 <sup>***</sup> (0,002)	0,029 <sup>***</sup> (0,003)	0,028 <sup>***</sup> (0,003)
Близость материального положения	-0,012 (0,008)		-0,003 (0,005)		-0,004 (0,005)	
Константа	-0,016 (0,033)	-0,042 (0,027)	0,024 <sup>***</sup> (0,008)	0,024 <sup>***</sup> (0,006)		
Фиксированные эффекты	да	да	нет	нет	нет	нет
Число наблюдений	6480	9506	6480	9506	6480	9506
R <sup>2</sup>	0,129	0,117	0,103	0,091		
Псевдо R <sup>2</sup>					0,198	0,188

В скобках указаны стандартные погрешности

\*  $p < 0,1$ , \*\*  $p < 0,05$ , \*\*\*  $p < 0,01$

Результаты свидетельствуют о том, наиболее сильный по влиянию оказался фактор соседства. Обучение в одной группе увеличивает вероятность возникновения дружбы, в зависимости от спецификации модели и факультета, от 10% до 28%. Влияние проживания в одном общежитии гораздо слабее – от 1% до 6%. При этом для дружеских связей студентов на факультете экономики влияние данного фактора значимо только для линейной модели с фиксированными эффектами и одной спецификации логистической модели и только на 10-ти процентном уровне значимости. Возможно, полученные оценки слабого влияния совместного проживания в общежитии на вероятность возникновения дружбы связаны со структурой выборки, с незначительной долей студентов, проживающих в общежитии в исследуемой совокупности. Переменные «проживание в одном общежитии» и «обучение в одной группе» отражают эффект соседства в студенческих социальных сетях. Полученные результаты позволяют сделать вывод о влиянии географической близости на формирование дружеских взаимодействий в социальных сетях студентов НИУ ВШЭ в Нижнем Новгороде.

Совпадение пола студентов статистически значимо влияет на вероятность возникновения дружбы во всех спецификациях и приводит к росту вероятности связи на

2,3-6,6%. Близость успеваемости – другая значимая характеристика. Для дружеской сети на факультете менеджмента приближение среднего балла на 1 пункт увеличивает вероятность возникновения дружбы между студентами на 1,7%-3,5%, на факультете экономики – 2,8%-3,4%. Для студентов факультета бизнес информатики воздействие близости среднего балла немного ниже (1,8%-2,3%), но также статистически значимо во всех спецификациях модели. Близость показателей материального положения в большинстве регрессий статистически незначима, однако следует отметить, что данный показатель измеряется достаточно грубо и субъективно. Таким образом, подтвердились предположения о том, что в социальных сетях наблюдается тенденция к формированию связей между студентами близкими по определенным характеристикам. Таким образом, результаты исследования подтверждают гипотезу о влиянии эффекта соседства и гомофилии на формирование дружеских связей.

### *3. Заключение*

Механизм образования связей по принципу близости индивидуальных черт характерен для различных типов социальных сетей. В результате личные социальные контакты людей происходят в значительной степени с людьми, близкими по социально-демографическим и личностным характеристикам. Это накладывает ограничения на получаемую от других людей информацию, формирование взглядов и мировоззрения (Mcpherson et al. 2001).

В данной работе анализировалось, как некоторые легко наблюдаемые факторы влияют на формирование дружеских социальных связей в студенческой среде. Наибольшим образом увеличивает вероятность возникновения дружеских связей «географическая» близость - принадлежность студентов к одной группе и совместное проживание в общежитии. Дружеские отношения ощутимо чаще возникают среди студентов одного пола. Кроме того значимым оказывается близость среднего балла: студенты с близкой успеваемостью будут дружить с большей вероятностью, чем студенты, чьи средние баллы различны. Полученные качественные результаты являются устойчивыми: они повторяются для сетей трех различных факультетов и различных спецификаций модели – линейной вероятностной модели с фиксированными эффектами и без фиксированных эффектов, логистической модели.

Количественная оценка роли разных факторов важна для понимания механизма социальных связей. Круг общения студента важен для его успеваемости: друзья участвуют в формировании отношения к учебе и своими успехами способны стимулировать собственное усердие, а также оказывать прямую поддержку, объясняя учебный материал и помогая в выполнении заданий. Хотя выбор друзей, несомненно, является добровольным, мы видим, что на вероятность её формирования влияют контролируемые факторы. Поэтому администрация вуза может целенаправленно формировать студенческие группы и расселять студентов в общежитии, например, чтобы оптимизировать общую успеваемость, или создать благоприятную среду для самых способных студентов, либо поддержать недостаточно подготовленных учащихся.