

# Что в профиле тебе моем данные «ВКонтакте» как инструмент изучения интересов современных подростков

**К. Н. Поливанова, И. Б. Смирнов**

Статья поступила  
в редакцию  
в декабре 2016 г.

**Поливанова Катерина Николаевна**  
доктор психологических наук, профессор, директор Центра исследований современного детства Института образования Национального исследовательского университета «Высшая школа экономики». E-mail: kpolivanova@hse.ru

**Смирнов Иван Борисович**

младший научный сотрудник Института образования Национального исследовательского университета «Высшая школа экономики». E-mail: ibsmirnov@hse.ru

Адрес: 101000, Москва, ул. Мясницкая, 20.

**Аннотация.** Интересы играют ключевую роль в психическом развитии ребенка. Однако их изучение связано с серьезными методологическими проблемами. Традиционно они исследуются с помощью опросных методов, не способных адекватно отразить многообразный и динамичный мир интересов взрослеющего человека. Авторы предлагают использовать для изучения интересов подростков их подписки на сообщества в социальной сети «ВКонтакте». Отличительной особенностью данных «ВКонтакте» является не только их масштаб, но и отсутствие цензурированности ответов, ха-

рактерной для опросов. Возможности применения метода продемонстрированы на примере одной московской школы: 674 учащихся, подписанных на 20 203 различных сообщества. Обнаружено, что интересы подростков дифференцированы в зависимости от их пола, возраста и успеваемости. Сила этой дифференциации показана на расширенном наборе данных, содержащем информацию об интересах 290182 пользователей «ВКонтакте». Установлено, что подписки на сообщества позволяют с большой точностью предсказать не только пол подростка (97%) и его возраст (98%), но и образовательные результаты школы, в которой он обучается (83%). Результаты исследования указывают на неоднородность нормы возраста, в частности на ее связь с полом подростка и его успеваемостью. Признание разнородности интересов и разнообразия нормы создает условия для гуманизации образования, в противном случае оно все более отчуждается от реального человека, игнорируя его собственный интерес.

**Ключевые слова:** социальные сети, «ВКонтакте», подростковый возраст, интересы, машинное обучение.

**DOI:** 10.17323/1814-9545-2017-2-134-152

Публикация подготовлена в рамках под- держанного РГНФ научного проекта № 16-06-00916.

Смена интересов, их развитие составляют, по Л. С. Выготскому, сущность подросткового возраста. Глава «Развитие интересов в переходном возрасте» в «Педагогике подростка» открывается часто цитируемой фразой: «Ключом ко всей проблеме

психологического развития подростка является проблема интересов в переходном возрасте» [Выготский, 1984а. С. 6]. И далее: «Отправной точкой научного исследования в этой области является признание того, что развиваются не только навыки и психологические функции ребенка (внимание, память, мышление и т. п.) — в основе психического развития лежит прежде всего эволюция поведения и *интересов ребенка*, изменение структуры направленности его поведения» (курсив наш. — К. П., И. С.) [Там же]. Сущность возраста в отечественной науке определяется, по Выготскому, как социальная ситуация развития, как отношение к «окружающей действительности, прежде всего социальной» [Выготский, 1984b. С. 258]. То есть отношение (во всяком случае применительно к подростковому возрасту) и интерес — если и не синонимы, то близкие понятия. Отношение в подростковом возрасте приобретает форму интереса. Следовательно, подростковый возраст может быть описан через структуру интересов.

Второе важное для нас базовое допущение Л. С. Выготского — единообразие нормы возраста. И у Выготского, и в других классических работах по возрастной периодизации — Пиаже, Фрейда, Эриксона — речь идет о возрасте как универсальной характеристике. В сущности, вся психология развития базируется на идее, метафоре «лестницы возрастов», единой для всех. Отклонение от этого универсального пути взросления трактуется как опережение или отставание, но никак не возможность разнообразия. Э. Бурман критикует эту позицию: «Нормальный ребенок — идеальный тип, созданный на основе результатов эмпирического исследования разных возрастов, — является поэтому вымыслом или мифом» [Бурман, 2006. С. 30]. Представление о норме, по Бурман, довлеет над реальными проявлениями индивидуальности.

Соединяя два тезиса — об интересе как сущности возраста и об универсальности нормы возраста, — мы формулируем вопрос: в какой мере интересы одной возрастной группы, а именно подростков, одинаковы, направлены на однотипные объекты, в какой мере эти объекты могут быть определены как единообразные, а если нет, то свидетельствует ли это об отклонении от нормального развития или о различных типах развития?

Традиционно интересы и связанные с ними конструкты — потребность, мотив, влечение, направленность — исследуются методом опроса, как правило, это бланковые методики-анкеты [Лубовский, 2003]. При их разработке в психологии и социальных науках всегда (за исключением проективных методов) исходно задается нормативный образ изучаемого явления, функции, процесса — объекта изучения. Он операционализируется, а за-

**Методы исследования интересов**

тем (как правило, в процессе апробации) задаются критерии соответствия-несоответствия реальных значений нормативным (средним). При индивидуальной диагностике оценивается соответствие эмпирически полученных значений нормативным. При этом, с одной стороны, задаваемые вопросы могут быть нерелевантны для индивида, у него может отсутствовать какое-либо отношение к тому, о чем его спрашивают, и тогда появляются случайные, неинформативные ответы [Bertrand, Mullainathan, 2001]. С другой стороны, действительно значимое для индивида может оставаться вне поля зрения исследователя. Особенно актуальной эта проблема становится применительно к исследованию интереса как значимой интимной сферы личности.

Альтернативой может стать проективный метод [Соколова, 1980], который базируется на фрейдовских представлениях о вынесении вовне внутренних переживаний личности (проекции). В основе методических приемов лежит предъявление испытуемому стимульного материала, который допускает бесконечное множество трактовок-интерпретаций; трактовка, которую дает испытуемый, рассматривается как связанная с уникальной, свойственной данной личности системой внутренних переживаний. Исследование поведения подростков в социальной сети можно рассматривать как изучение неосознанных проекций индивида. Сам выбор контента (из практически бесконечного множества) является проекцией собственного интереса, проявляет его как для самого индивида, так и для исследователя, если такие выборы удастся эксплицировать.

Независимо от выбранных методов — качественных, количественных, объективных, проективных и т. д. — существует и еще одна проблема, особенно сильно искажающая потенциальные результаты при исследовании подросткового возраста: цензурированность (социальная желательность) ответов. Как только исследование затрагивает действительно важную для респондента тему, существует большая опасность ухода и маскировки, не всегда осознаваемых даже самим респондентом. Применительно к подростку это особенно значимо, поскольку одной из характеристик возраста является появление собственного «пространства», которое намеренно скрывается от внешнего наблюдателя [Осорина, 2011].

Таким образом, теоретически мы фокусируемся на ключевой характеристике — интересе, а наблюдению и эмпирическому изучению она недоступна. Возникает настоятельная необходимость в новых методах, которые обнаружили бы подростковые интересы *per se*.

Мы полагаем, что шагом в направлении создания таких методов может быть изучение поведения подростков в виртуальном пространстве. Социальные сети становятся для современных подростков естественной средой, школьники пользуются мо-

бильными устройствами постоянно [Королева, 2016а]. Создавая свой профиль, подросток выбирает друзей, контент, подписывается на группы. Все это он делает неподконтрольно, по собственному выбору, причем выбирая из почти бесконечного числа возможностей. Мы исходим из того, что подростковые профили представляют своеобразную картину интересов пользователя, которая дает уникальную возможность их объективного изучения.

Постоянная проверка новостей и сообщений в социальных сетях стала ритуалом, свойственным подростковой повседневности [Королева, 2016б]. При этом подростки отдают предпочтение социальной сети «ВКонтакте»: в качестве основной ее называют 86% респондентов [Королева, 2016а]. Новостная лента «ВКонтакте» формируется из контента, размещаемого друзьями пользователя, а также из информации от групп (сообществ), на которые пользователь подписан. Сообщества в социальных сетях предоставляют уникальную возможность изучать интересы подростков, эти сведения недоступны традиционным эмпирическим исследованиям [Lewis et al., 2008]. Всего на «ВКонтакте» более 26 млн сообществ, на самые популярные из них подписаны сотни тысяч пользователей. Сообщества могут быть посвящены любимой компьютерной игре, фильму, книге, актрисе, музыкальному исполнителю или политике. Большинство организаций и брендов — от Мариинского театра до тату-салона — имеют свое сообщество на «ВКонтакте». Любые интересы и хобби, от вышивания до квантовой физики, представлены в социальной сети. Практически у каждой школы есть свое сообщество «Подслушано в...», в котором школьники делятся слухами. Сообщества могут быть посвящены сексуальным отношениям, неразделенной любви, определенному образу жизни или настроению («Плед, какао и камин»).

Таким образом, контент социальной сети можно рассматривать как стимульный материал для проекции интересов пользователя, а выявленные подписки — как картину его индивидуальных интересов.

Мы использовали набор данных, содержащий информацию о 674 учащихся одной московской школы: об их успеваемости (средний балл за последний год), поле, параллели, в которой они учатся (с 5-го по 10-й класс), а также о списке сообществ в социальной сети «ВКонтакте», на которые они подписаны. Общее число сообществ, на которые подписан хотя бы один учащийся, составило 20 203. Мы исключили из рассмотрения сообщества, на которые подписаны менее 10 учащихся, в итоге для анализа использовалась информация о 883 сообществах.

### **Сообщества «ВКонтакте»**

### **Данные и методология исследования**

Для сбора данных было использовано специальное программное обеспечение (программа), выполняющее запросы к API (публичному интерфейсу приложения) «ВКонтакте». Программное обеспечение запускается представителем образовательной организации, подающим на вход список имен учащихся. С помощью этого списка программа сопоставляет учащихся с их профилями на «ВКонтакте». Для увеличения охвата пользователей программа осуществляет поиск не только среди пользователей, указавших, что они учатся в заданной образовательной организации, но и среди их друзей в социальной сети. Помимо этого программа использует обширный словарь альтернативных форм имени (Анна, Аня, Анечка, Анютка и т. п.). После выполнения процедуры сопоставления и загрузки информации программа удаляет все имена, фамилии и идентификаторы «ВКонтакте». Обезличенные данные передаются для дальнейшего исследования. Эффективность такого подхода, а также отсутствие значительных смещений полученной выборки были продемонстрированы в работе [Смирнов, Сивак, Козьмина, 2016].

Чтобы получить представление о многообразии интересов учащихся одной школы, мы составили карты их интересов. Для каждого из сообществ была вычислена доля девочек/мальчиков среди его подписчиков, а также средний балл его подписчиков и их средний возраст. В зависимости от полученных значений названия групп были расположены на координатной плоскости. Ось  $X$  соответствует возрасту учащихся (выраженному через номер параллели, в которой они учатся), ось  $Y$  — успеваемости школьников.

**Карты интересов** На рис. 1 представлены сообщества, в которых среди подписчиков больше 50% составляют девочки, а на рис. 2 — сообщества с долей мальчиков среди подписчиков больше 50%.

Уже эта качественная информация дает основания для некоторых рассуждений и оценок. Сообщества, на которые в основном подписаны мальчики, расположились по диагонали от верхнего левого угла к нижнему правому, т. е. от области высоких оценок у шестиклассников к области низких оценок у девятиклассников. У девочек карта интересов в этом отношении структурирована слабее. Такое расположение сообществ связано с естественной тенденцией оценок к снижению от 5-го класса к 10-му, причем у мальчиков она выражена сильнее.

На представленных картах довольно отчетливо видно, что состав сообществ меняется от младших классов к старшим. Девочки в целом позже, чем мальчики, подписываются на сообщества (меньше точек в левой части рисунка). Плохо успевающие девочки (правый нижний угол) подписаны на такие сообщества, как «Любовный Гороскоп», «Необычный Гороскоп», «Тысяча чертей,

Рис. 1. Карта интересов учащихся. Сообщества с долей девочек среди подписчиков больше 50%

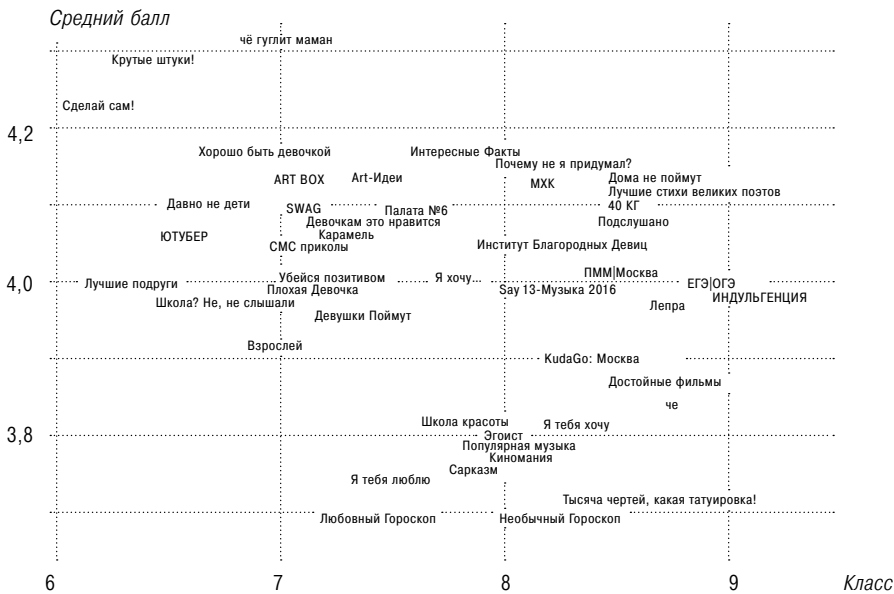


Рис. 2. Карта интересов учащихся. Сообщества с долей мальчиков среди подписчиков больше 50%

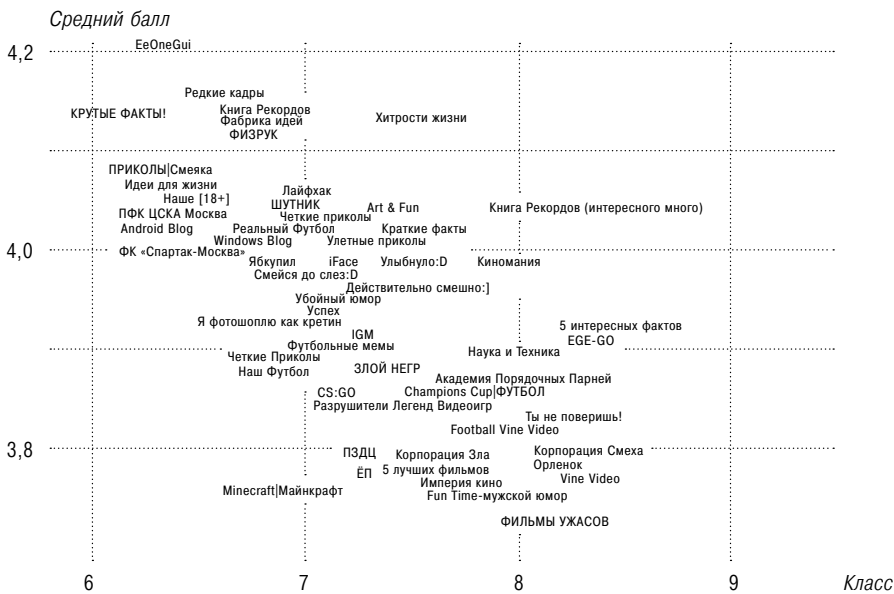


Таблица 1. Сообщества, в которых доля девочек превышает 80%

Доля девочек, %	Название группы	Категория
100	Институт Благородных Девиц	Девичье
100	Лучшие подруги	Девичье
100	Хорошо быть девочкой	Девичье
97	Девушки Поймут	Девичье
96	Карамель	Девичье с фантазиями о семье, верности и т. д.
95	40 КГ	Похудение, здоровое питание и спорт
95	Девочкам это нравится	Девичье
93	Дома не поймут	Девичье, взросление
91	Школа красоты	Красота
90	#SWAG	Девичье, слэщавое, юмор
87	ИДЕИ для творчества	Рукоделие и т. п.
85	Я хочу...	Прочее
84	Взрослей	Любовь, мечты, банальности
83	Лучшие стихи великих поэтов	Стихи, преимущественно о любви
82	Сделай сам!	Рукоделие
82	Плохая Девочка	Девичье с уклоном в особенность и стервозность
82	Фабрика идей	Рукоделие и т. п.
82	Арт-Идеи	Рукоделие и т. п.
80	Давно не дети	Любовь
80	Задыхаюсь	Идеальная любовь, исключительность

какая татуировка!». У их лучше успевающих одноклассниц появляются сообщества, связанные с учебой, например «ЕГЭ/ОГЭ». И еще более высокие оценки имеют девочки, подписанные на сообщество «40 КГ» — о здоровом питании и фигуре, о спорте. При этом уже в 6-м классе у девочек появляется сообщество «Давно не дети» — о сексе, причем довольно откровенном и с весьма грубой лексикой. Мальчиков этого возраста на это сообщество подписано меньше, чем девочек. «Мужской юмор Fun Time» у мальчиков появляется только в 8-м классе, причем у плохо успевающих. У мальчиков с более высокой успеваемостью также есть группа, посвященная ЕГЭ и ОГЭ, как и у их одноклассниц.

### Дифференциация по полу и успеваемости

В табл. 1 и 2 представлены сообщества с наибольшей долей девочек и наибольшей долей мальчиков среди подписчиков,

Таблица 2. Сообщества, в которых доля мальчиков превышает 80%

Доля мальчиков,%	Название группы	Категория
100	CS: GO	Компьютерные игры
100	Windows Blog	Компьютеры
100	Игрофан	Компьютерные игры
96	Разрушители Легенд Видеоигр	Компьютерные игры
96	Football Vine Video	Футбол
95	Minecraft   Майнкрафт	Компьютерные игры
93	Champions Cup   ФУТБОЛ	Футбол
92	IGM	Компьютерные игры
91	Футбольные мемы	Футбол
91	Успех	Прочее
91	Реальный Футбол	Футбол
90	ФУТБОЛЬНЫЕ МЕМЫ	Футбол
88	Академия Порядочных Парней	Мальчиковое
86	Наш Футбол	Футбол
82	Книга рекордов (интересного много)	Интересные факты
80	ЗЛОЙ НЕГР	Расистские шутки
80	ФИЛЬМЫ УЖАСОВ	Фильмы ужасов

а в табл. 3 и 4 — с наибольшим и наименьшим средним баллом среди подписчиков. Содержание выделенных таким образом сообществ было изучено более подробно (не менее 15 последних записей). Им были присвоены категории, раскрывающие основное содержание публикуемых записей. Часть групп не удалось классифицировать в силу большой разнородности размещенной в них информации, такие группы были отнесены к категории «прочее». Все изученные группы включают большое число картинок, видеофрагментов. Текстовые записи короткие и, как правило, являются комментариями к изображениям или видео, а собственно тексты единичны.

Наиболее популярными являются сообщества, посвященные романтическим отношениям (включая сексуальные), компьютерным играм, футболу и различным «идеям» — примерам изобретений или неожиданному использованию привычных вещей. Практически все записи, за исключением сугубо романтических, содержат шутки и «приколы». Хотя есть и собственно группы «Юмор».



Таблица 3. Сообщества со средним баллом подписчиков выше 4,1

Средний балл	Название группы	Категория
4,30*	Чёгуглитмаман	Юмор, определения себя
4,28*	Крутые штуки!	Идеи
4,22*	Сделай сам!	Идеи
4,20*	ЕеOneGuy	Прочее
4,16*	Хорошо быть девочкой	Девичье
4,15*	Интересные Факты	Интересные факты
4,15*	Редкие кадры	Редкие кадры
4,14*	Почему не я придумал?	Идеи
4,13*	Книга Рекордов	Интересные факты
4,13*	Крутые факты!	Интересные факты
4,12	Хитрости жизни	Идеи
4,12	Дома не поймут	Юмор, определения себя
4,12	Арт-Идеи	Идеи
4,12	МХК	Юмор
4,12	Фабрика идей	Идеи
4,12	ФИЗРУК	Прочее
4,11	Фабрика идей	Идеи
4,11	ART BOX	Рукоделие и т. п.
4,10	Лучшие стихи великих поэтов	Стихи

\* Значение значимо выше среднего на уровне 0,05.

В текстах часто встречается обценная лексика (иногда с заменой одного знака в слове или в сокращении), хотя в сообществах, посвященным романтическим отношениям, много таких слов, как «милота», «обнимашки» и т. п., и умильных картинок.

Интересы мальчиков фокусируются вокруг футбола, компьютерных игр и разного рода юмора. Часто используется обценная лексика и сленг. В группах, на которые подписаны мальчики, много фотографий признанных секс-символов. Довольно много в этих группах, независимо от содержания и основной направленности, видеофрагментов или текстов с неожиданной концовкой, т. е. построенных по структуре анекдота, много неожиданных ситуаций, например видеофрагмент с многочисленными столкновениями автомобилей на скользкой дороге («Скользкая Канада»).

Таблица 4. Сообщества со средним баллом подписчиков ниже 3,8

Средний балл	Название группы	Категория
3,68*	Любовный Гороскоп	Гороскопы
3,68*	Необычный Гороскоп	Гороскопы
3,70*	Тысяча чертей, какая татуировка!	Татуировки
3,71*	ФИЛЬМЫ УЖАСОВ	Фильмы
3,74*	Я тебя люблю	Романтика
3,75*	Сарказм	Юмор
3,76*	Fun Time — мужской юмор	Юмор
3,76*	Minecraft   Майнкрафт	Компьютерные игры
3,77*	Киномания	Фильмы
3,77*	Vine Video	Видео
3,77*	Империя Кино	Фильмы
3,77*	ЁП	Юмор
3,78*	Популярная музыка	Музыка
3,78*	5 лучших фильмов	Фильмы
3,79	Корпорация Зла	Юмор
3,79	Эгоист	Прочее
3,79	Орленок	Юмор
3,79	ПЗДЦ	Юмор
3,79	Корпорация Смеха	Юмор

\* Значение значимо ниже среднего на уровне 0,05.

Карта сообществ, на которые подписаны девочки, совсем другая. Девочек интересуют собственно девичьи темы, в первую очередь романтика. Много постов «о нем» — он должен понять, принять, он обязательно вернется и пожалеет о разрыве даже спустя годы. Много про разрыв отношений, переживание одиночества, покинутости. Можно предположить, что интерес вызывает именно сюжетный узел, а бесконфликтная благополучная ситуация такой драматургией не обладает. Много сообществ и записей про отношения с близкой подругой, это описанная еще Фрейдом фаза подростковости, когда возникает сильно эротически окрашенное чувство к другу своего пола. Группа «Плохая Девочка» также про девичьи темы, но с элементами язвительности и цинизма, что и следует из названия: образ героини — умная, сильная, не слишком заботящаяся о том, что про нее думают окружающие. Довольно много групп посвящено рукоделию,

их можно, хотя и с натяжкой, поставить в соответствие с группами для мальчиков «Фабрика идей». В лексике девичьих групп меньше вульгаризмов и обсценных выражений, хотя говорить что-то определенное о соотношении числа сниженных речевых оборотов у мальчиков и девочек пока преждевременно. При этом много уменьшительно-ласкательных оборотов, таких как «милота», «обнимашки».

Если группы, на которые подписаны мальчики, структурно можно связать с энергичным действием, то девичьи группы — про переживания.

Сообщества «отличников» открывает «чёгуглитмаман» — очень своеобразная группа, в которой представлены вымышленные запросы матери в форме «Окейгугл, почему моя дочь..?» Например, «Окейгугл, почему доча ждет письмо из какого-то Хогвартса?». То есть подростки описывают воображаемые беспокойства матери (вероятно, обоих родителей) по поводу того, что происходит с дочерью. Это довольно сложная когнитивная конструкция: подросток представляет себе, что в его поведении может быть непонятным родителям, притом что ему самому это представляется очевидным и естественным. Именно очевидность для одних (самих подростков) и непонятность для других (родителей) делает каждый пост смешным (в большей или меньшей степени). В этой же группе встречаются упоминания социальной сети «Одноклассники» и почты в домене mail.ru как примеры чего-то устаревшего. Но юмор этот не злой и не грубый. Отчасти к этой же категории можно отнести и сообщество «Дома не поймут», но это, скорее, самоописания, которые, как явствует из названия группы, непонятны старшим, — вспомним «личный миф» (*personal fable*), характеристику, введенную Д. Элкиндом [Alberts, Elkind, Ginsberg, 2007].

Кроме того, в списке сообществ, на которые подписаны «отличники», есть и «Фабрика идей» — забавные или серьезные неожиданные варианты использования привычных вещей, странные их сочетания и т. д., есть юмор, фильмы, музыка и «Лучшие стихи великих поэтов» (преимущественно о любви). Секс также представлен практически во всех сообществах, но, как правило, внутри каких-то других тем (например, неожиданное возвращение родителей домой, когда молодая пара занимается сексом).

Сообщества школьников с более низким средним баллом успеваемости составляют совсем иную подборку. Это гороскопы («любовный» и «необычный»), затем «Тысяча чертей, какая татуировка!» — множество фотографий татуировок, иногда смешные картинки с татуировками. Также есть подборки фильмов, видеоклипов, музыка, много юмора, подчас весьма грубого.

В записях из сообществ обеих категорий пользователей — и «отличников», и «двоечников» — выражения встречаются весьма разнообразные, в том числе вульгаризмы и сниженная лексика.

Подростковые сообщества практически не содержат школьной тематики. Школа, учение, содержание школьных предметов, таким образом, исключены из сферы интересов детей (только предстоящие экзамены попадают в поле их интересов, вероятно, по совершенно внешним причинам, по необходимости). Слабо-успевающие школьники при этом не интересуются и экзаменами.

Таким образом, качественное сравнение групп, на которые подписаны мальчики и девочки, «отличники» и «двоечники», выявляет различия в интересах между этими категориями подростков. Различия существуют и между подростками разных возрастов. Чтобы точно охарактеризовать эти различия, требуется дальнейшая работа с материалом. Вероятнее всего, адекватным методом сравнения будет контент-анализ. Он не входил в задачи настоящего этапа исследования ввиду большого количества сообществ (более 800).

Карты интересов позволяют качественно оценить дифференциацию интересов подростков в зависимости от пола, возраста и успеваемости. При этом остается открытым вопрос, насколько значимы обнаруженные различия. Найти ответ на этот вопрос традиционными методами представляется затруднительным. Поясним это на примере. Предположим, мы бы хотели оценить степень дифференциации успеваемости по полу. При традиционном подходе пол был бы назван независимой переменной, а успеваемость — зависимой. Была бы построена регрессионная модель, предсказывающая успеваемость по полу и ряду контрольных переменных. Величина коэффициента перед переменной пола и уровень его значимости послужили бы основанием для вывода о силе связи между успеваемостью и полом. Несмотря на широкое распространение, такой подход вызывает серьезную критику [Berk, 2004], начиная с классической работы Лимера [Leamer, 1983], и не является единственно возможным [Breiman, 2001]. Очевидно, что для ответа на интересующий нас вопрос традиционный подход непригоден, так как интересы пользователей — переменная огромной размерности (сотни тысяч) и она не может быть использована как зависимая переменная в регрессионной модели.

### **Степень дифференциации интересов**

Мы утверждаем, что степень дифференциации интересов можно оценить через их предсказательную силу, а именно через точность модели, предсказывающей на основании информации об интересах подростков их принадлежность к интересующей нас группе (мальчики/девочки, младшеклассники/старшеклассники и т. п.). То есть если возможно построить модель, предсказывающую с высокой точностью пол подростков, — значит, их ин-

### **Предсказательная сила интересов**

тересы дифференцированы по полу. Как и в случае с проверкой статистических гипотез, обратное не обязательно верно: если не удалось построить такую модель — это не значит, что дифференциация отсутствует.

Поясним эту мысль примером. Рассмотрим соответствие между полом и внешностью человека. Разумеется, именно пол определяет внешность, а не наоборот. Однако требовать предсказать внешность по полу было бы неразумно, так как переменная «пол» может принимать всего два значения, а переменная «внешность» — миллиарды различных значений. Тем не менее не вызывает сомнений существование соответствия между полом и внешностью. Не вызывает сомнений как раз потому, что по внешности мы можем предсказать пол с большой точностью. То же самое может быть верно и для интересов: если интересы человека позволяют предсказать его пол с той же точностью, что и внешность, — значит, интересы не менее полоспецифичны.

Построение предсказательной модели требует значительного большего объема данных, чем в используемом нами наборе, поэтому мы дополнительно собрали информацию обо всех пользователях «ВКонтакте», указавших, что они учатся или учились в одной из школ Санкт-Петербурга, и родившихся в период с 1993 по 2002 г. Про каждого пользователя были известны пол, год рождения, средний балл ЕГЭ выпускников его школы за последние пять лет и список подписок на сообщества. Всего этот городской набор содержал информацию о 290 182 пользователях. Общее число различных сообществ, на которые подписан хотя бы один учащийся, — 886 191. Так как в этом случае не было возможности сопоставить данные «ВКонтакте» с реальными, для повышения достоверности набора данных из него были исключены те пользователи, у которых не было друзей в социальной сети из указанной ими школы.

### **Построение модели**

С учетом вышесказанного не имеет принципиального значения, какую именно предсказательную модель использовать. Мы отдаем предпочтение подходу, предложенному М. Косински с соавторами [Kosinski, Stillwell, Graepel, 2013]. Во-первых, потому, что это наиболее известная из работ, выполненных на данных, аналогичных нашим, — а значит, можно будет сопоставить российские результаты с мировыми. Во-вторых, потому, что предложенный М. Косински подход прямолинеен, и для его понимания не требуется владеть продвинутыми методами машинного обучения. Аналогичный подход к анализу больших данных используется и в социальных исследованиях [Eagle, Pentland, 2009].

Из городского набора данных мы исключили пользователей, подписанных менее чем на 50 сообществ, и сообщества, на которые были подписаны менее 50 пользователей. Результирующий

набор данных содержал 116 912 пользователей и 40 774 сообщества. Участие в каждом из сообществ было закодировано с помощью бинарных переменных  $a_j$  ( $j = 1, \dots, 40\,774$ ), где  $a_j = 0$ , если пользователь не подписан на сообщество  $j$ , и  $a_j = 1$ , если пользователь подписан на сообщество  $j$ . Весь набор данных таким образом представляет собой матрицу  $116\,912 \times 40\,774$ ,  $(i, j)$ -элемент которой равен 1, если пользователь с номером  $i$  подписан на сообщество с номером  $j$ , и равен 0 в противном случае. Затем мы использовали сингулярное разложение этой матрицы для выделения 100 главных компонент  $b_k$  ( $k = 1, \dots, 100$ ), характеризующих интересы пользователей.

Переменные  $b_k$  использовались в качестве предикторов в логистической регрессии. Целевыми переменными выступали пол пользователя, его принадлежность к определенной возрастной группе и факт обучения в школе с наилучшими/наихудшими образовательными результатами. Чтобы избежать переобучения моделей, была использована перекрестная проверка с разбиением на десять частей.

Построенные нами модели на городском наборе данных позволяют предсказать пол пользователя с точностью 97%. Они также позволяют отнести пользователя к одной из двух возрастных групп с точностью 98%, если разница между ними составляет 9 лет (2002 и 1993 год рождения). Если разница между возрастными группами составляет 4 года (2002 и 1998 год рождения), то точность падает до 88%. Модель позволяет различить даже разницу в 2 года: точность составляет 70%. Модель позволяет разделить пользователей из 1% наиболее успевающих школ и 1% наименее успевающих школ с точностью 83%. Отнести пользователей к 50% наиболее успевающих школ или к 50% наименее успевающих школ модель позволяет с точностью 62%. Такая низкая точность неудивительна, так как индивидуальные достижения школьников из этих двух категорий школ в значительной степени пересекаются между собой.

По интересам подростков мы можем предсказать их пол с точностью, близкой к 100%. Это означает, что если мы возьмем набор интересов некоей девочки, то мы вполне можем встретить похожий набор интересов у другой девочки, но никогда ни у одного мальчика. Теоретически высокую точность предсказания можно обеспечить за счет простого запоминания того, какому набору интересов какой пол должен соответствовать. В этом случае мы не могли бы сделать содержательных выводов из высокой точности предсказания. Для того чтобы избежать подобной ситуации, мы использовали перекрестную проверку (кросс-валидация). В этом случае модель строится на основании одних данных, а ее точность проверяется на других.

**Предсказательная сила подписок на сообщества**

**Заключение** Виртуальная среда, в частности социальная сеть «ВКонтакте», которой пользуются подавляющее большинство российских школьников, содержит помимо индивидуальных профилей-аккаунтов еще и сообщества: публичные страницы (паблики) и группы. Они собирают пользователей по их интересам. Мы рассматривали факт подписки на сообщество как индикатор интереса школьника к контенту, который превалирует в данном сообществе. На основании имеющихся в нашем распоряжении данных по одной московской школе (списки учащихся и их средний балл) удалось идентифицировать и затем сгруппировать сообщества, на которые подписаны школьники. Полученные результаты в целом не противоречат имеющейся информации об интересах школьников. Так, они согласуются с представлениями о доминантах — доминантах дали, усилия и романтики и эгоцентрической установке, — описанных в «Педагогике подростка» Л. С. Выготского, а также об эгоцентрической направленности у Д. Элkinда [Elkind, 1967].

Использованные математические процедуры дали возможность существенно расширить и уточнить эти представления, а главное — наполнить их конкретным содержанием посредством анализа контента групп. Для более точного и полного описания и квалификации содержания, которым наполняются паблики, требуется разработать специальную процедуру. Мы полагаем, что адекватным формализованным инструментом сравнения содержания групп будет контент-анализ. Однако прямое применение имеющихся версий этого инструмента пока невозможно, поскольку контент групп включает не только текстовые записи — их как раз меньшинство, — но и видеоматериалы, и картинки, и музыку. Поэтому на данном этапе мы ограничились качественным описанием содержания групп с уточнением лексики, которую используют пользователи.

Таким образом, мы представляем новый инструмент анализа интересов подростков — наиболее информативной стороны их жизни. Результаты применения нашего инструмента в целом не противоречат имеющимся в литературе представлениям, однако существенно их уточняют и расширяют. Наиболее важным из полученных результатов мы считаем возможность дифференциации интересов школьников по их успеваемости. Мы можем теперь обоснованно выдвигать гипотезу о том, что развитие интересов школьников происходит по разным траекториям в зависимости от их успеваемости. То есть информация о разнообразии образовательных траекторий подтверждается более точной информацией о внутренних психологических процессах подростка — его интересах.

Проведенное исследование имеет как теоретическую, так и сугубо практическую значимость. Теоретически оно указывает на неоднородность нормы возраста, на ее связь с полом

подростка и его успеваемостью. Если разнородность интересов мальчиков и девочек так или иначе подразумевалась (хотя никак не учитывалась в образовании), то различие интересов детей одного пола, связанное с успеваемостью, «обсуждалось» вне научного поля, например в кино и литературе о детях. Теперь эта диверсификация может быть изучена предметно и конкретно. Очень важным направлением дальнейших исследований мы считаем более точное описание дифференцированных групп, выделенных не по одному основанию (мальчики/девочки, «отличники»/«двоечники»), а по нескольким, как минимум по трем: полу, возрасту/классу, успеваемости. Также напрашивается дифференциация подростков по социально-экономическому и культурному бэкграунду семей, но эта информация недоступна в массовом масштабе, так как не представлена в аккаунтах пользователей.

Практическая значимость информации о характерных интересах подростков различных групп нам представляется очевидной. Во-первых, наличие такой информации создает условия для гуманизации образования, в противном случае оно все более отчуждается от реального человека, игнорируя его собственный интерес; наличие такой информации создает условия для развития образовательно-досуговой деятельности. Во-вторых, признание разнородности интересов и разнообразия нормы может служить основанием для практической работы с подростками, базой для развития их интересов, для узко направленной психологической поддержки.

Система образования продолжает ориентироваться на идею универсальности нормы, широко представленную в работах по психологии развития. Эта идея, идущая еще от педологии, сегодня, как нам представляется, все меньше соответствует действительности.

1. Бурман Э. (2006) Деконструктивная психология развития. Ижевск: Удмуртский университет, ERGO.
2. Выготский Л. С. (1984а) Педология подростка // Выготский Л. С. Собрание сочинений: в 6 т. Т. 4. Детская психология. М.: Педагогика. С. 6–242.
3. Выготский Л. С. (1984b) Проблема возраста // Выготский Л. С. Собрание сочинений: в 6 т. Т. 4. Детская психология. М.: Педагогика. С. 244–269.
4. Королева Д. О. (2016а) Всегда онлайн: использование мобильных технологий и социальных сетей современными подростками дома и в школе // Вопросы образования / Educational Studies Moscow. № 1. С. 205–224. doi: 10.17323/1814-9545-2016-1-205-224
5. Королева Д. О. (2016b) Исследование повседневности современных подростков: присутствие в социальных сетях как неотъемлемая составляющая общения // Современная зарубежная психология. Т. 5. № 2. С. 55–61.

## Литература



6. Лубовский Д. В. (2003) Психодиагностические методы в работе с учащимися 3–4-х классов. Москва; Воронеж: Флинта, Московский психолого-социальный институт.
7. Миллер С. (2002) Психология развития: методы исследования. СПб.: Питер.
8. Осорина М. В. (2011) Секретный мир детей в пространстве мира взрослых. СПб.: Питер.
9. Соколова Е. Т. (1980) Проективные методы исследования личности. М.: Изд-во Московского ун-та.
10. Смирнов И. Б., Сивак Е. В., Козьмина Я. Я. (2016) В поисках утраченных профилей: достоверность данных «ВКонтакте» и их значение для исследований образования // Вопросы образования / Educational Studies Moscow. № 4. С. 106–122. doi: 10.17323/1814-9545-2016-4-106-122
11. Alberts A., Elkind D., Ginsberg S. (2007) The Personal Fable and Risk-Taking in Early Adolescence // Journal of Youth and Adolescence. Vol. 36. No 1. P. 71–76.
12. Berk R. A. (2004) Regression Analysis: A Constructive Critique. Vol. 11. Advanced Quantitative Techniques in the Social Sciences. Thousand Oaks, CA: Sage.
13. Bertrand M., Mullainathan S. (2001) Do People Mean What They Say? Implications for Subjective Survey Data // The American Economic Review. Vol. 91. No 2. P. 67–72.
14. Breiman L. (2001) Statistical Modeling: The Two Cultures // Statistical Science. Vol. 16. No 3. P. 199–231.
15. Eagle N., Pentland A. S. (2009) Eigenbehaviors: Identifying Structure in Routine // Behavioral Ecology and Sociobiology. Vol. 63. No 7. P. 1057–1066.
16. Elkind D. (1967) Egocentrism in Adolescence // Child Development. Vol. 38. No 4. P. 1025–1034.
17. Kosinski M., Stillwell D., Graepel T. (2013) Private Traits and Attributes Are Predictable from Digital Records of Human Behavior // Proceedings of the National Academy of Sciences. Vol. 110. No 15. P. 5802–5805.
18. Leamer E. E. (1983) Let's Take the Con out of Econometrics // The American Economic Review. Vol. 73. No 1. P. 31–43.
19. Lewis K., Kaufman J., Gonzalez M., Wimmer A., Christakis N. (2008) Tastes, Ties, and Time: A New Social Network Dataset Using Facebook. com. // Social Networks. Vol. 30. No 4. P. 330–342.

## What's in My Profile: VKontakte Data as a Tool for Studying the Interests of Modern Teenagers

**Katerina Polivanova**

Doctor of Sciences in Psychology, Professor, Director of the Center for Modern Childhood Research, Institute of Education, National Research University Higher School of Economics. E-mail: kpolivanova@hse.ru

Authors

**Ivan Smirnov**

Junior Researcher at the Institute of Education, National Research University Higher School of Economics. E-mail: ibsmirnov@hse.ru

Address: 20 Myasnikskaya St., 101000 Moscow, Russian Federation.

Children's interests play a key role in their psychological development. However, research in this field is associated with serious methodological problems, as it has traditionally used questionnaire surveys that cannot adequately describe the diverse and dynamic world of interests of a developing person. The article suggests using the information on VKontakte communities followed by teenagers, in order to explore their interests. Apart from being comprehensive, VKontakte data is, unlike questionnaire answers, also uncensored. The method's potential demonstrated through the example of a Moscow school with 674 students following 20,203 various VKontakte communities. It reveals that teenagers' interests vary depending on their gender, age, and academic performance. The degree of such variance is demonstrated on an extended set of data on the interests of 290,182 VKontakte users. It transpires that communities followed by teenagers predict with high accuracy not only their gender (97%) and age (98%) but also the performance of the schools they attend (83%). The findings point to the heterogeneity of age-related behavior patterns, in particular to their correlation with gender and academic achievements. Acknowledgement of the heterogeneity of interests and the diversity of age-related behavior patterns creates conditions for the further development of student-centered education, in the absence of which education is becoming more and more alienated from real life, ignoring the interests of real people.

Abstract

social networking services, VKontakte, adolescence, interests, machine learning.

Keywords

- Alberts A., Elkind D., Ginsberg S. (2007) The Personal Fable and Risk-Taking in Early Adolescence. *Journal of Youth and Adolescence*, vol. 36, no 1, pp. 71–76.
- Berk R. A. (2004) Advanced Quantitative Techniques in the Social Sciences. *Regression Analysis: A Constructive Critique*, vol. 11. Thousand Oaks, CA: Sage.
- Bertrand M., Mullainathan S. (2001) Do People Mean What They Say? Implications for Subjective Survey Data. *The American Economic Review*, vol. 91, no 2, pp. 67–72.
- Breiman L. (2001) Statistical Modeling: The Two Cultures. *Statistical Science*, vol. 16, no 3, pp. 199–231.
- Burman E. (2006) *Dekonstruktivnaya psikhologiya razvitiya* [Deconstructing Developmental Psychology]. Izhevsk: Udmurt State University, ERGO.
- Eagle N., Pentland A. S. (2009) Eigenbehaviors: Identifying Structure in Routine. *Behavioral Ecology and Sociobiology*, vol. 63, no 7, pp. 1057–1066.
- Elkind D. (1967) Egocentrism in Adolescence. *Child Development*, vol. 38, no 4, pp. 1025–1034.

References

- Koroleva D. (2016a) Vsegda onlayn: ispolzovanie mobilnykh tekhnologiy i sotsialnykh setey sovremennymi podrostkami doma i v shkole [Always Online: Using Mobile Technology and Social Media at Home and at School by Modern Teenagers]. *Voprosy obrazovaniya / Educational Studies Moscow*, no 1, pp. 206–224. doi: 10.17323/1814-9545-2016-1-205-224
- Koroleva D. (2016b) Issledovanie povsednevnosti sovremennykh podrostkov: prisutstvie v sotsialnykh setyakh kak neotyemlemaya sostavlyayushchaya obshcheniya [A Study of the Daily Life of Modern Teenagers: The Presence in Social Networks as an Integral Component of Communication]. *Journal of Modern Foreign Psychology*, vol. 5, no 2, pp. 55–61. doi:10.17759/jmfp.201605020
- Kosinski M., Stillwell D., Graepel T. (2013) Private Traits and Attributes Are Predictable from Digital Records of Human Behavior. *Proceedings of the National Academy of Sciences*, vol. 110, no 15, pp. 5802–5805.
- Leamer E. E. (1983) Let's Take the Con out of Econometrics. *The American Economic Review*, vol. 73, no 1, pp. 31–43.
- Lewis K., Kaufman J., Gonzalez M., Wimmer A., Christakis N. (2008) Tastes, Ties, and Time: A New Social Network Dataset Using Facebook. com. *Social Networks*, vol. 30, no 4, pp. 330–342.
- Lubovsky D. (2003) *Psikhodiagnosticheskie metody v rabote s uchashchimisya 3–4 klassov* [Psychodiagnostic Methods in Teaching Students of Grades 3–4]. Moscow; Voronezh: Flinta, Moscow Psychological and Social Institute.
- Miller S. (2002) *Psikhologiya razvitiya: metody issledovaniya* [Developmental Research Methods]. Saint Petersburg: Piter.
- Osorina M. (2011) *Sekretny mir detey v prostranstve mira vzroslykh* [The Secret World of Children in the Adult Environment]. Saint Petersburg: Piter.
- Smirnov I., Sivak E., Kozmina Y. (2016) V poiskakh utrachennykh profiley: dostovernost' dannykh "VKontakte" i ikh znachenie dlya issledovaniy obrazovaniya [In Search of Lost Profiles: The Reliability of VKontakte Data and Its Importance for Educational Research]. *Voprosy obrazovaniya / Educational Studies Moscow*, no 4, pp. 106–122. doi: 10.17323/1814-9545-2016-4-106-122
- Sokolova E. (1980) *Proektivnye metody issledovaniya lichnosti* [Projective Techniques in Personality Research]. Moscow: Moscow State University.
- Vygotsky L. S. (1984a) Pedologiya podrostka [Adolescent Pedology]. *Sobranie Sochineniy L. S. Vygotskogo* [Collected Works of L. S. Vygotsky], vol. 4. Moscow: Pedagogika, pp. 5–242.
- Vygotsky L. (1984b) Problema vozrasta [The Problem of Age]. *Sobranie Sochineniy L. S. Vygotskogo* [Collected Works of L. S. Vygotsky], vol. 4. Moscow: Pedagogika, pp. 244–269.