

СЕТЕВОЙ АНАЛИЗ ВЗАИМОДЕЙСТВИЙ В ОНЛАЙН СООБЩЕСТВАХ «ЖИВОГО ЖУРНАЛА»

Введение

«Живой Журнал» (livejournal.com) является самым популярным сервисом блогов в России. О его популярности и влиятельности можно судить по возрастающему числу блогов политиков, количеству упоминаний в прессе, а также по скандалам и дискуссиям в масс-медиа, которые были инициированы или опосредованы данным сервисом.

В связи с этим некоторые исследователи видят в Интернете и блогосфере признаки «публичной сферы», описанной Юргеном Хабермасом, однако сам автор полагает иначе. С его точки зрения, Интернет не только не способствует становлению «публичной сферы», но и мешает ей, так как разделяет единое дискуссионное пространство на множество разрозненных групп по интересам, в которых узкий круг участников обсуждает ряд локальных и специфических проблем.

Предметом исследования является скандал на социологическом факультете МГУ, начавшийся в феврале 2007 г. с записи в «Живом Журнале» о процедурных нарушениях на выборах декана, по результатам которых им вновь стал В.И. Добренков. Автором данной записи была студентка социологического факультета, которая вскоре перевелась в ГУ ВШЭ. Запись вызвала широкую дискуссию в блогосфере, которая продолжилась в марте, когда неформальная студенческая организация «OD-Group» выдвинула ряд претензий к руководству социологического факультета, среди которых были указания на завышенные цены в столовой, сдачу учебных аудиторий в аренду, запрет на чтение некоторых западных авторов и пропаганда Православной социологии. Следующим витком скандала стало обнаружение плагиата в учебниках и книгах В.И. Добренкова и его сотрудников. Финальным аккордом стало сообщение о том, что ряд активистов из «OD-Group» отчислены в марте 2008 г.

Таким образом, за год скандал вышел за рамки обсуждения одной записи в блоге, став достоянием общественности посредством блогосферы и онлайн СМИ. О степени резонанса можно судить хотя бы по тому, что о данном скандале было написано в «New York Times» незадолго после его начала.

Методология

В связи с этим необходимо иное теоретическое обоснование коммуникации в Интернете и «Живом Журнале» в частности. В качестве таковой мы предлагаем сетевой подход к взаимодействиям в рамках той или иной коммуникационной платформы, будь то блог, форум, «социальная сеть» и т.д. В основе нашего подхода лежат три послылки.

Публичность. Первая заключается в том, что исследователь, выступая в качестве стороннего наблюдателя, может использовать только те формы взаимодействия, которые явлены ему непосредственно и зафиксированы в текстовой или иной мультимедийной форме средствами технологической платформы, на которой происходит исследуемое взаимодействие.

Специфичность. Согласно второй посылке, при анализе компьютерно-опосредованной коммуникации (computer-mediated communication) необходимо учитывать специфику медиасреды, в которой эта коммуникация происходит. Это также означает, что у каждой технологической платформы существуют уникальные черты, требующие концептуализации в рамках теоретической и математической модели взаимодействия. То есть методы сбора и анализа данных, разработанные для «Живого Журнала», невозможно будет полностью воспроизвести применительно к, скажем, «Одноклассники.ру».

Сетевая структура взаимодействия. Последней теоретической посылкой нашей работы является использование сетевой методологии при анализе взаимодействия в «Живом Журнале». Таким образом, при анализе первичную роль для нас будут играть структурные эффекты, оказываемые тем или иным видом взаимодействия.

Определение взаимодействия в «Живом Журнале»

Согласно первой и второй теоретическим посылкам нашей работы, мы разделяем взаимодействие в «Живом Журнале» на 4 типа: 1) Добавление в друзья («френды»); 2) Ссылки/цитирование; 3) Комментирование; 4) Написание записей в блог («постинг»).

Каждый из представленных типов может быть описан и представлен в виде сети взаимодействий. Для анализа текстов сообщений в блогах также применяется особый вид сетевого анализа текстов – «map analysis», разработанный К. Карли и реализованный в виде комплекса программного обеспечения Automap

и ORA. Сетевой анализ коммуникации каждого типа направлен на выявление структурных регулярностей на уровне диад (например, реципрокность), триад (например, транзитивность), а также всей сети (например, кластеризация). Для проверки статистической значимости обнаруженных структурных эффектов использованы ERGM (Exponential Random Graph Models или p^*) модели.

Формальные сети «френдов». Первый тип является наиболее формальным и менее всего интерпретируемым в связи с тем, что интенция данного действия скрыта для исследователя. Один блоггер может добавить к себе в френд-ленту другого блоггера для демонстрации дружеских отношений, из соображений реципрокности, для публичной демонстрации связи, если добавляемый блоггер является известным и/или популярным, для оперативного оповещения о новых записях и т.д. С точки зрения внешнего наблюдателя, каким является исследователь, все эти мотивы совершенно неразличимы.

В данной работе исследование строится на анализе наблюдаемого акта «френдинга» – добавления блоггером авторов других дневников в собственную френд-ленту. Практика добавления в «друзья» имеет два измерения: социальное и технологическое. Интерпретация первого может быть затруднительной даже в том случае, когда мы имеем возможность обратиться к мнению самих авторов блога, добавившего/удалившего другой блог в «друзья». Связь авторов через «дружбу» в «Живом Журнале» вовсе не означает наличие тесных связей в реальной жизни: список друзей может как пересекаться со списком знакомых и друзей в реальности, так и не быть с ним связанным вовсе. Добавление другого дневника в «друзья» – это довольно значимый шаг, что связано с тремя обстоятельствами. Во-первых, автор добавленного блога получает доступ к приватным записям, доступ к которым обычно ограничивается списком друзей. Во-вторых, добавляя кого-либо в свои «друзья», автор подписывается на появление в своей френд-ленте записей авторов добавленных блогов. Внимание блоггера не безгранично, поэтому обычно авторы стараются ограничить френд-ленту допустимыми размерами, что зависит во многом от того, сколько времени автор готов тратить на чтение или просматривание записей своих «друзей». И, наконец, в-третьих, акт френдинга – это публичное действие, когда любой может увидеть список френдов любого блоггера. И хотя акт френдинга далеко не всегда означает наличие реальных связей между авторами, в тех случаях, когда авторы имеют непримиримые позиции по каким-либо вопросам или состоят в разных сегментах, не признающих легитимное существование друг друга, они едва ли предпочтут сделать видимым для остальных интерес друг к другу.

С технологическим определением дела обстоят несколько проще, так как функциональные характеристики акта добавления в друзья одинаковы для всех.

Добавляя кого-либо в «друзья», блоггер предоставляет автору добавленного блога доступ к своим приватным – «подзамочным» – записям. Таким образом, добавление в «друзья» служит косвенным признаком доверия к автору добавленного блога. Однако существуют сложности, связанные с технологическими особенностями платформы. В частности с ее политиками приватности, в которых существует три уровня: 1) публичный (записи доступны для всех); 2) приватный (для всех «друзей»); 3) конфиденциальный (записи видны только самому автору блога). Второй уровень приватности может быть дифференцирован – пользователь может выбирать группы «френдов», которым будет открыт доступ к конкретной записи. В связи с этим стороннему наблюдателю невозможно различить, какая из политик приватности в данном блоге. В ситуации, когда достоверные данные о практиках приватности стороннему наблюдателю получить невозможно, то остается рассматривать эту практику как косвенный признак доверия.

Источником данных для данной сети послужил список блоггеров, которые являются членами специального сообщества «Живого Журнала» «*socio-log*». В базу данных были отобраны все участники сообщества – 1210 блоггеров. Итоговая матрица состояла из 1210 столбцов на 1210 строк, если участник сообщества добавил другого участника сообщества в «друзья», то в матрице фиксировалось наличие связи.

Сети гиперссылок. Второй тип связей между блоггерами, представленный сетями гиперссылок, является более информативным, так как позволяет судить, какую информацию блоггер считает интересной, важной и достойной распространения. Имея данные о том, на какие внешние по отношению к «Живому Журналу» ресурсы ссылается тот или иной блоггер, мы можем выяснить, на какие ресурсы ссылаются различные стороны конфликта, вычислить самые популярные ресурсы и блоггеров, наиболее часто ссылающихся на внешние источники. Также собранные данные позволяют нам построить бимодальную сеть (*bimodal network*), которая позволяет оценить, какие блоггеры связаны между собой посредством ссылок на одинаковые источники и какие онлайн-ресурсы связаны между собой за счет того, что на них ссылаются одни и те же блоггеры.

Бимодальная сеть строится на основе общей матрицы, в строках которой находятся записи блогов, а в столбцах – статьи на внешних сайтах. Из нее можно получить две взвешенные матрицы смежности: записи в блогах, связанные между собой ссылками на одинаковые источники и источники, связанные между собой на основе количества записей в блогах, ссылающихся на них. Однако взаимосвязь записей и статей довольно сложно интерпретировать, поэтому мы сократили размерность обеих матриц за счет того, что перешли от записей к блоггерам и от статей к сайтам. Таким образом, мы получили две взвешенные

матрицы, в одной из которых отображены связи между блоггерами, ссылающимися на одинаковые ресурсы, а в другой – сайты, на которые ссылаются одинаковые пользователи. Вес связи в каждой из матриц обозначает количество постов блоггера/статей сайта, которыми связаны сайты и блоги соответственно.

Сети взаимного комментирования. Третий тип взаимодействия в «Живом Журнале» позволяет измерять не только структуру актуальных взаимодействий с другими блоггерами, но также ее интенсивность. Комментирование чужих записей и ответ на чужие комментарии к собственным является одной из основных практик ведения блога в «Живом Журнале». Собственно она-то и является наиболее интерактивным элементом взаимодействия как там, так и во всей блогосфере в целом. Таким образом, выяснив, какие блоггеры чаще всего комментируют друг друга, мы можем судить о том, кто, каким образом и с какой интенсивностью связаны друг с другом.

Данные собраны во взвешенную матрицу смежности между блоггерами, оставившими хотя бы один комментарий, где веса представляют количество раз, когда блоггер из строки ответил на запись или комментарий блоггера из столбца.

Постинг как взаимодействие. Наконец, написание записей в блог является самым насыщенным для интерпретации, так как именно записи наиболее полно представляют точку зрения блоггера по той или иной проблеме.

Статистическое моделирование с помощью ERGM

Основной целью применения ERGM-моделей являлась попытка определить наличие структурных «сил» или «эффектов» в сети связей между блогами участников сообщества «sociolog». Под структурными эффектами мы понимаем устойчивые эндогенные факторы, влияющие на то или иное взаимодействие. Наиболее понятным примером такого фактора может служить «реципрокность», лежащая в основе сетей обмена подарками, неформальной поддержки и т.д. Другим примером структурного эффекта может служить «транзитивность» в сетях дружбы: «друг моего друга – мой друг». Обратный эффект – интранзитивность – лег в основу теоретической модели сильных/слабых связей М. Грановеттера в виде «запрещенной триады» (*forbidden triad*), смысл которой заключается в том, что не может быть такой триады, в которой две пары акторов были бы связаны между собой, а третья – нет, ибо это не имеет смысла. Если, конечно, один из участников не ведет двойную жизнь и не является шпионом под прикрытием.

Обнаружение подобных структурных эффектов позволит выявить основные паттерны, которые позволяют описать, как организовано внимание участников сообщества. Мы сможем ответить на вопросы типа: Характерно ли для участников сообщества оказывать взаимное внимание? Или свойственно ли данной сети образовывать «звезд», чьи дневники читают многие, но сами они не склонны оказывать внимание?

Для решения такой задачи элементарной описательной статистики в виде плотности сети, центральных и подсчета «треугольников» не достаточно. Поэтому мы воспользовались семейством моделей, основанных на экспоненциальных случайных графах (Exponential Random Graph Models, ERGM), также известных в литературе как ρ^* -модели. Суть данных моделей заключается в том, чтобы подобрать статистический параметр для каждого из интересующих структурных эффектов (например, реципрокность, транзитивность, цикличность) и с его помощью построить модель, которая наилучшим образом описывала бы имеющиеся данные. Иными словами, мы определяем наблюдаемую сеть как одну из множества возможных сетей с таким же количеством акторов, которые генерируются неким стохастическим процессом, и пытаемся понять, какие локальные структурные эффекты привели к образованию данной сети. Например, мы можем предположить, что наблюдаемая сеть дружбы может быть образована за счет таких локальных структурных эффектов, как реципрокность, транзитивность, склонность к централизации (образованию «звезд» – узлов, обладающих большим числом связей, нежели остальные) или кластеризации. Каждому структурному эффекту соответствует определенная конфигурация узлов и связей, ниже мы опишем их более подробно.

Оценка качества модели изначально производилась с помощью метода максимального правдоподобия (MLE, Maximum Likelihood Estimation), однако ввиду сложности вычислений, от него пришлось отказаться и перейти к оценке с помощью псевдоправдоподобия (*pseudolikelihood*) и симуляции методом Монте-Карло с помощью сетей Маркова (Markov Chain Monte Carlo, MCMC).

В общем виде модель выглядит следующим образом:

$$P(Y = y) = \left(\frac{1}{k}\right) \exp\left\{\sum_A \theta_A g_A(y)\right\},$$

где y – наблюдаемая сеть; Y – вероятностное распределение сетей; $g_A(y)$ – статистика структурного эффекта в наблюдаемой сети; θ_A – статистический параметр данного эффекта, который нам необходимо оценить; k – нормализующая константа; A – все возможное количество графов, которое может получиться при заданном числе узлов.

Анализ данных

В идеале исследователи пытаются подобрать модель, которая включала бы в себя как можно больше параметров, дабы наилучшим образом объяснить процесс формирования интересующих их сетей. Однако на практике зачастую так и не получается оценить некоторые из параметров. Например, в одной из работ авторы, анализирувавшие сеть дружеских отношений среди 1681 школьников, обнаружили, что для ее описания достаточно транзитивности типа АТ-Т и таких атрибутов, как класс и раса. В частности, подтвердилась поговорка «друг моих друзей – мой друг» и выяснилось, что склонность к формированию связей внутри одного класса сильнее, чем внутри расы, хотя оба эффекта сильны и имеют статистическую значимость. Другой проблемой использования r^* -моделей является их дегенерация, проявляющаяся в несходимости модели. Так, даже значимые параметры могут приводить к дегенеративным моделям.

Мы столкнулись с обеими проблемами, указанными выше. Из 22 моделей, параметры которых мы пытались оценить, только в одной, самой элементарной, они оказались значимыми и не привели к дегенерации. Эта была модель общей склонности к образованию связей между участниками сообщества «sociology». Значение параметра составило «-5,13», что означает логарифмическое отношение шансов (*log-odds*) к образованию связи между любыми двумя актерами составляет «-5,13», т.е. связи между ними, скорее всего, не будет.

При попытках усложнить модель, добавив туда новые параметры, нам удавалось добиться большего правдоподобия (AIC), однако вместе с этим модели становились дегенеративными. Это не обязательно означает, что данные параметры ухудшают модель. Можно обратиться к настройкам симуляций MCMC (Markov Chain Monte Carlo), увеличив для них объем выборки, количество итераций или длину цепи Маркова, а также наложить дополнительные условия на симуляцию графов, например, контролируя их плотность или распределение степеней. Однако эти манипуляции значительно увеличивают время оценки каждой из моделей, особенно на слабых компьютерах.

В связи с этим мы решили сначала подобрать оптимальный набор параметров модели, а затем бороться с ее дегенеративностью.

Из табл. 1 видно, как по мере добавления новых параметров улучшается правдоподобие модели (значение AIC уменьшается), а также повышается значимость параметров. Так, в модели 4 транзитивность стала значимой на 5-процентном уровне, а значение ее параметра указывает на небольшую склонность в сторону транзитивности, т.е. возможно подтверждение поговорки «друг моих друзей – мой друг».

Таблица 1.

Модель	Эффект	Параметр	Значение	Значимость	Правдоподобие (AIC)
1	Связи	arc	-5,28	0	84227
2	Связи	arc	-5,26927	0	84226
	Реципрокность	reciprocity	-0,03942	1	
3	Связи	arc	-5,27704	0	84196
	Транзитивность	AT-T	0,15248	0,1	
4	Связи	arc	-5,132362	0	84011
	Транзитивность	AT-T	0,17959	0,05	
	Интранзитивность	A2P-T	-0,031484	0	

Таким образом, получение качественной модели зависит от времени и ресурсов компьютера. Однако уже на данном этапе мы можем видеть, что значимыми параметрами являются отсутствие стремления к образованию связей и небольшая склонность к транзитивности. Это в контексте наших данных говорит о том, что блоггеры, участвующие в сообществе «sociology», не стремятся добавлять друг друга в друзья, однако в том случае, когда связи между блоггерами присутствовали, они склонны к эффекту транзитивности, т.е. правило «друг моего друга – мой друг» оказалось справедливым для данного типа отношений.

Сеть гиперссылок. Всего в данной бимодальной сети было 189 записей в блогах, содержащих 628 ссылок на различные онлайн-ресурсы. В связи с тем, что при таком количестве узлов весьма сложно получить наглядную визуализацию связей, в большинстве случаев мы ограничимся табличной формой представления наиболее популярных статей, сайтов и блоггеров.

Из табл. 2 видно, что абсолютным лидером по количеству ссылок стала веб-страница студенческой организации «OD-Group». Это вполне предсказуемый результат, так как на данном сайте выкладывались материалы, дававшие впоследствии повод для дискуссии в блогосфере. Среди них были фотографии, нормативные акты, ссылки на публикации в СМИ, интервью и т.д.

Остальные девять позиций занимают страницы из раздела «Нам пишут» веб-страницы социологического факультета МГУ. В настоящий момент все эти записи недоступны, поэтому весьма сложно интерпретировать столь высокую их популярность.

Таблица 2.

Web-страницы, на которые чаще всего ссылались блоггеры

№	Статья	Число ссылок
1	http://www.od-group.org/	126.000
2	http://www.socio.msu.ru/?s=life&p=for_us&t=19	56.000
3	http://www.socio.msu.ru/?s=life&p=for_us&t=14	56.000
4	http://www.socio.msu.ru/?s=life&p=for_us&t=15	56.000
5	http://www.socio.msu.ru/?s=life&p=for_us&t=16	56.000
6	http://www.socio.msu.ru/?s=life&p=for_us&t=17	56.000
7	http://www.socio.msu.ru/?s=life&p=for_us&t=18	56.000
8	http://www.socio.msu.ru/?s=life&p=for_us&t=24	56.000
9	http://www.socio.msu.ru/?s=life&p=for_us&t=20	56.000
10	http://www.socio.msu.ru/?s=life&p=for_us&t=21	56.000

В табл. 3 представлены ресурсы, на которые пришлось наибольшее количество ссылок. Первое место сайта «OD-Group» здесь аналогично результатам по веб-страницам. Однако официальный сайт социологического факультета оказался на четвертой позиции, в отличие от предыдущих результатов. Его обошли такие порталы как «Полит.ру» и «Газета.ру». Такие результаты могут свидетельствовать о том, что несмотря на высокую цитируемость отдельных страниц с официального сайта социологического факультета, на данных порталах было по-прежнему больше статей с небольшими показателями количества ссылок. Остальные позиции занимают новостные и аналитические ресурсы либерального толка, а на десятом месте – сервис для загрузки и скачивания файлов.

Таблица 3.

Ресурсы, на которые чаще всего ссылались блоггеры

№	Сайт	Число ссылок
1	od-group.org	159.000
2	polit.ru	62.000
3	gazeta.ru	50.000
4	socio.msu.ru	46.000
5	echo.msk.ru	28.000
6	izbrannoe.ru	23.000
7	city-fm.ru	21.000
8	svobodanews.ru	19.000
9	expert.ru	19.000
10	ifolder.ru	15.000

В табл. 4 представлены блоггеры, которые наиболее часто использовали ссылки на внешние ресурсы. Любопытно, что в отличие от веб-страниц и веб-сайтов, здесь практически не фигурируют блоггеры, занимавшие позицию декана Добренькова. Таким образом, можно сделать предварительный вывод о том, что в блогосфере формальные институции представлены весьма слабо и поэтому они уступают блоггерам либерального толка в возможностях обозначения собственной позиции.

Таблица 4. Блоггеры, которые наиболее часто ссылались на внешние источники

№	Пользователь	Число ссылок
1	olegivanov62	223.000
2	_falkon_	116.000
3	msu	107.000
4	3_gum	88.000
5	od-group	88.000
6	yashin	73.000
7	__temnota__	65.000
8	vd-effect	62.000
9	marchenk	50.000
10	rauwolfia	45.000

Сеть комментирования. На основе взаимного комментирования блоггерами друг друга можно построить сетевую диаграмму, которая достаточно четко показывает, что в данном виде взаимодействия можно выделить обособленные группы с одним лидером, которые взаимодействуют исключительно друг с другом. Среди таких лидеров мы можем выделить следующих блоггеров: ivan-gogh, kouprjanov, plestscheev, kreont, nataly_demina и marsky. По визуализации можно судить, что каждый из них собрал вокруг себя аудитории, которые практически не пересекаются между собой. Было бы логично предположить, что у этих подгрупп должно быть единое понимание ситуации. Данную гипотезу можно проверить с помощью текстового анализа контента сообщений.