Доклад о работах Jure Leskovec

Изменение графов во времени и выявление антисоциального поведения в интернете

Каюмов Эмиль

ММП ВМК МГУ

Спецсеминар
«Алгебра над алгоритмами и эвристический поиск закономерностей»

14 октября 2015

План

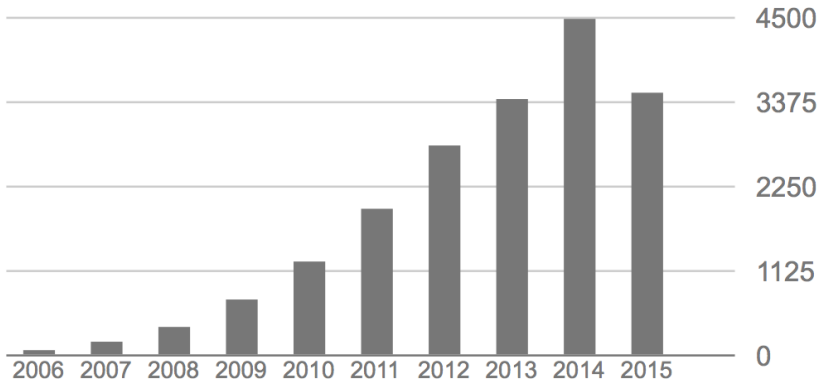
Содержание

Jure Leskovec

- 2004 – B.Sc. in Computer Science, University of Ljubljana, Slovenia
- 2008 – Ph.D. in Computational and Statistical Learning, Carnegie Mellon University, USA
- 2008-2009 – Postdoctoral researcher in Department of Computer Science, Cornell University, USA
- 2009-. . . – Assistant Professor, Stanford University, USA

Jure Leskovec

- Имеет более 100 публикаций
- Руководит научной группой
- Ведёт курс «Mining Massive Datasets» на Coursera

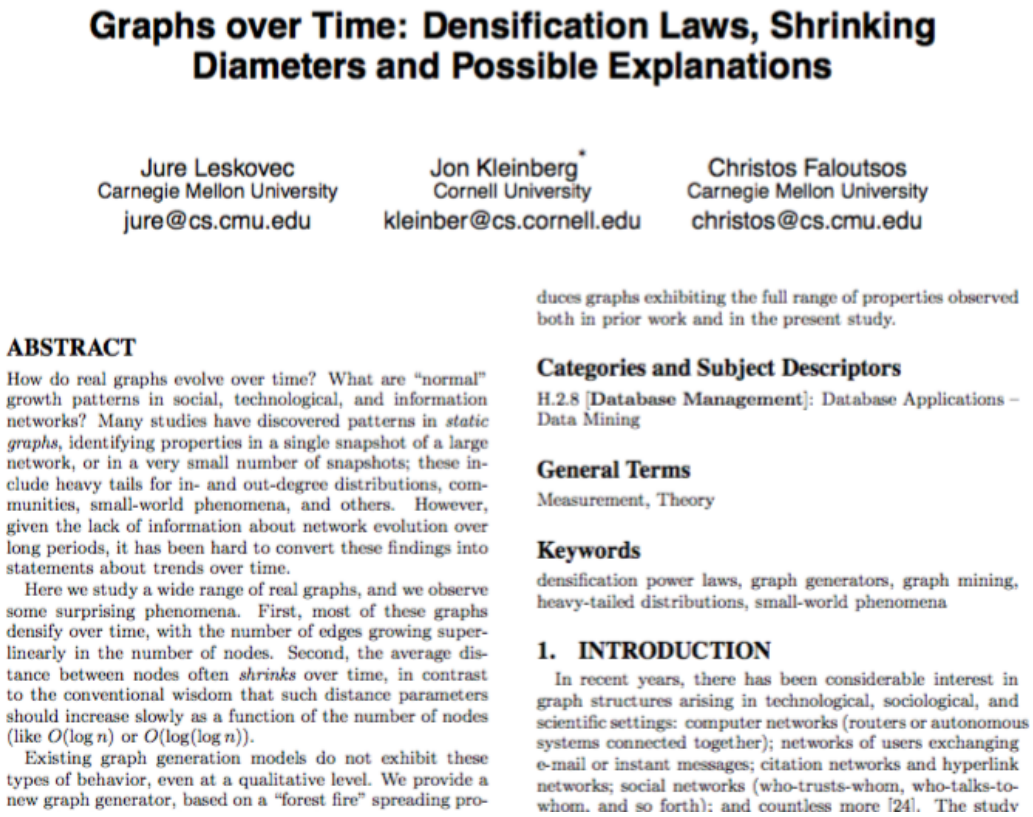


Количество ссылок за год

«My research focuses on mining and modeling large social and information networks, their evolution, and diffusion of information and influence over them. Problems I investigate are motivated by large scale data, the Web and on-line media.»

Содержание

Graphs over Time: Densification Laws, Shrinking Diameters and Possible Explanations



Graphs over Time: Densification Laws, Shrinking Diameters and Possible Explanations

Jure Leskovec
Carnegie Mellon University
jure@cs.cmu.edu

Jon Kleinberg*
Cornell University
kleinber@cs.cornell.edu

Christos Faloutsos
Carnegie Mellon University
christos@cs.cmu.edu

ABSTRACT

How do real graphs evolve over time? What are "normal" growth patterns in social, technological, and information networks? Many studies have discovered patterns in *static graphs*, identifying properties in a single snapshot of a large network, or in a very small number of snapshots; these include heavy tails for in- and out-degree distributions, communities, small-world phenomena, and others. However, given the lack of information about network evolution over long periods, it has been hard to convert these findings into statements about trends over time.

Here we study a wide range of real graphs, and we observe some surprising phenomena. First, most of these graphs densify over time, with the number of edges growing superlinearly in the number of nodes. Second, the average distance between nodes often *shrinks* over time, in contrast to the conventional wisdom that such distance parameters should increase slowly as a function of the number of nodes (like $O(\log n)$ or $O(\log(\log n))$.

Existing graph generation models do not exhibit these types of behavior, even at a qualitative level. We provide a new graph generator, based on a "forest fire" spreading process, that has a simple, intuitive justification, requires few parameters (like the "flammability" of the forest), and produces graphs exhibiting the full range of properties observed both in prior work and in the present study.

Categories and Subject Descriptors

H.2.8 [**Database Management**]: Database Applications – Data Mining

General Terms

Measurement, Theory

Keywords

densification power laws, graph generators, graph mining, heavy-tailed distributions, small-world phenomena

1. INTRODUCTION

In recent years, there has been considerable interest in graph structures arising in technological, sociological, and scientific settings: computer networks (routers or autonomous systems connected together); networks of users exchanging e-mail or instant messages; citation networks and hyperlink networks; social networks (who-trusts-whom, who-talks-to-whom, and so forth); and countless more [24]. The study

- Jure Leskovec, Jon Kleinberg, Christos Faloutsos.
- ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD), 2005. Best research paper award.

Предпосылки работы

- Какие законы распространяются на большинство реальных социальных графов?
- Как графы изменяются во времени?
- Можно ли сгененировать реальный социальный граф?

Зачем это может быть нужно?

- Генерация графов для исследований.
- Выделение подграфа для работы некоторых алгоритмов.
- Экстраполяция существующих графов.
- Обнаружение странного поведения в сетях.

Ранее

- В основном социальные графы изучались статичными.
- Основные идеи об изменении социальных графов:
  1. Среднея степень вершин остаётся постоянной (эквивалентно: количество рёбер растёт линейно с ростом количества вершин).
  2. Диаметр графа растёт медленно (как медленно растущая функция от размера графа).

- Степень вершины – количество входящих и исходящих из вершины рёбер.
- Диаметр – наибольшее расстояние между любыми парами вершин.

Предложение

- Сети уплотняются со временем.
- Количество рёбер растёт быстрее количества вершин $\Rightarrow$ средняя степень вершин увеличивается.

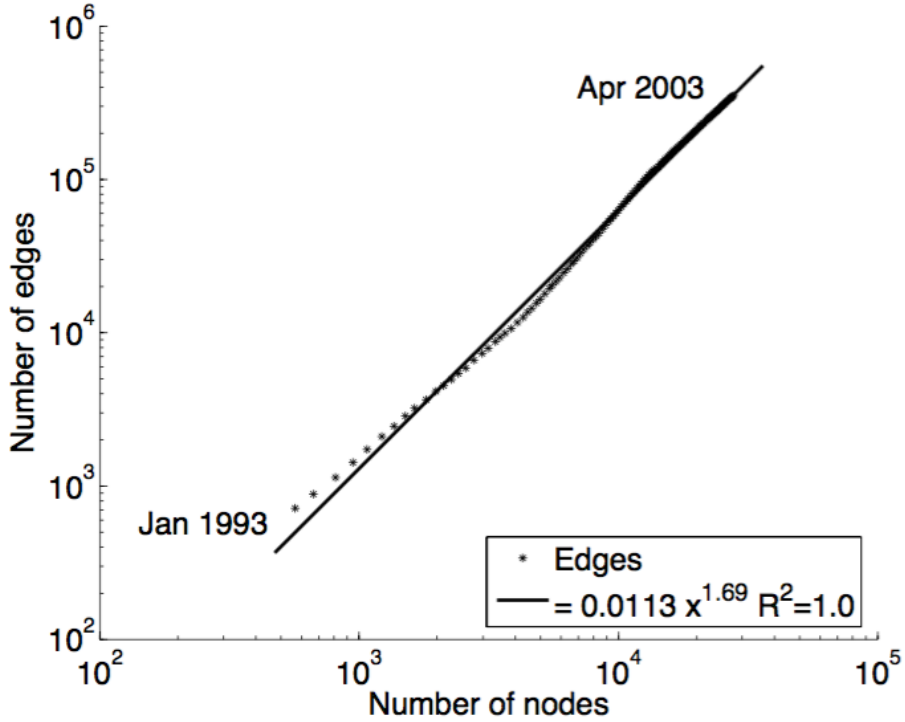
Закон увеличения средней степени вершин

$$e(t) \propto n(t)^{\alpha}$$

- $e(t)$ – количество рёбер, $n(t)$ – количество вершин.
- Экспонента сжатия: $1 \leq \alpha \leq 2$.
- $a = 1$ – линейный рост и постоянная средняя степень вершин.
- $a = 2$ – плотный граф, где в среднем сохраняется доля вершин, с которыми соединена каждая вершина.

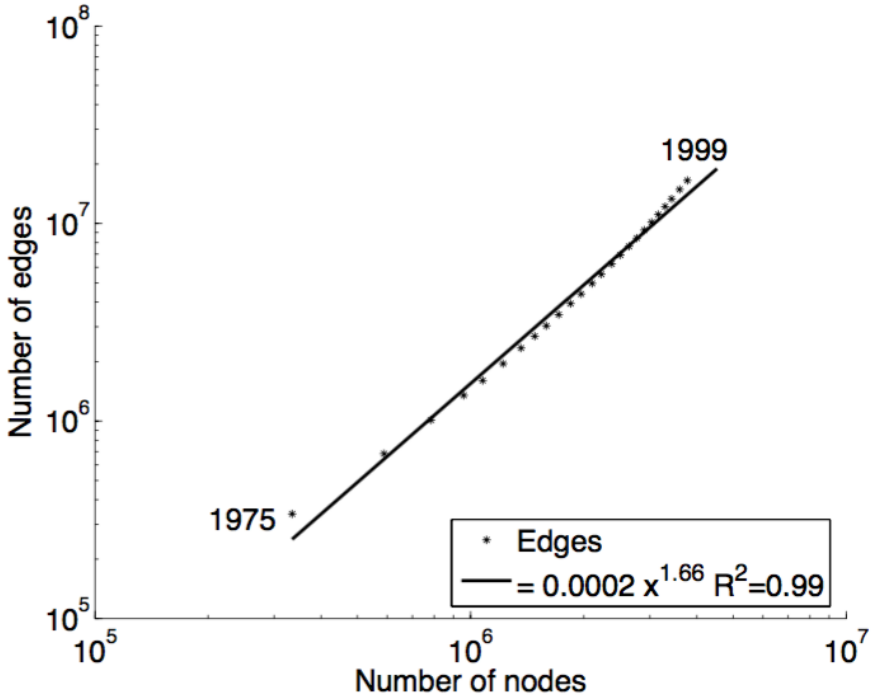
Реальный граф: цитирования в физике

- Цитирование среди статей по физике.
- Данные из arXiv.
- В 1992 году 1239 статей и 2717 цитирований.
- В 2003 году 29555 статей и 352807 цитирований.
- Каждая точка соответствует ситуации на определённый месяц.
- $\alpha = 1.69$.



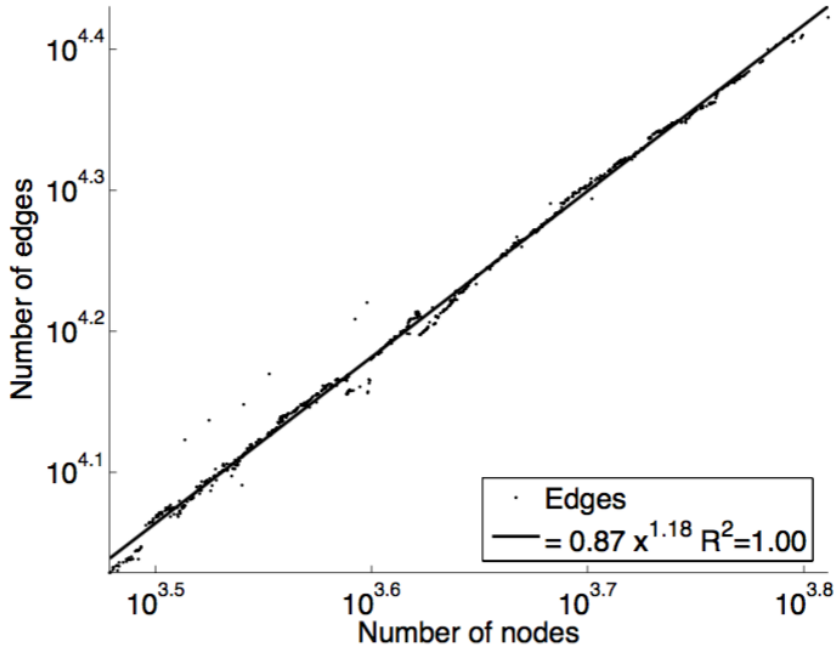
Реальный граф: цитирование в патентах

- Цитирование среди патентов в США.
- В 1975 году 334000 патентов и 676000 цитирований.
- В 1999 году 2.9 миллионов патентов и 16.5 миллионов цитирований.
- Каждая точка соответствует ситуации на определённый год.
- $\alpha = 1.66$.



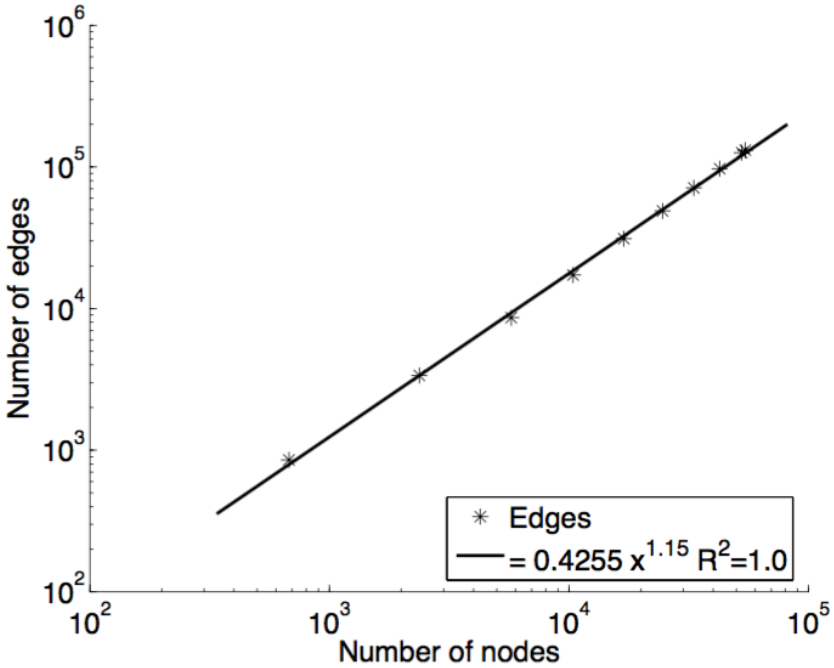
Реальный граф: граф интернета

- Обмен пакетами среди маршрутизаторов.
- Не только добавление вершин и рёбер, но и удаление.
- В 1997 году 3000 вершин и 10000 рёбер.
- В 2000 году 6000 вершин и 26000 рёбер.
- Каждая точка соответствует ситуации на каждый день.
- $\alpha = 1.18$.



Реальный граф: сооавторство

- Связь между авторами и статьями, в написании которых они участвовали.
- Данные из arXiv по нескольким категориям.
- В 1992 году 318 вершин и 272 рёбра.
- В 2000 году 58000 вершин (20000 авторов и 38000 статей) и 133000 рёбер.
- $\alpha = 1.15$.



Изменение диаметра

- Предыдущие работы говорили о том, что диаметр медленно увеличивается (как $O(\log N)$ или $O(\log \log N)$).
- Авторы говорят, что с ростом сети эффективный диаметр графа будет медленно уменьшаться в большинстве случаев.

- Диаметр – наибольшее расстояние между любыми парами вершин.
- Эффективный диаметр – минимальное расстояние, на котором 90% вершин достижимы друг из друга.

Эксперименты с эффективным диаметром

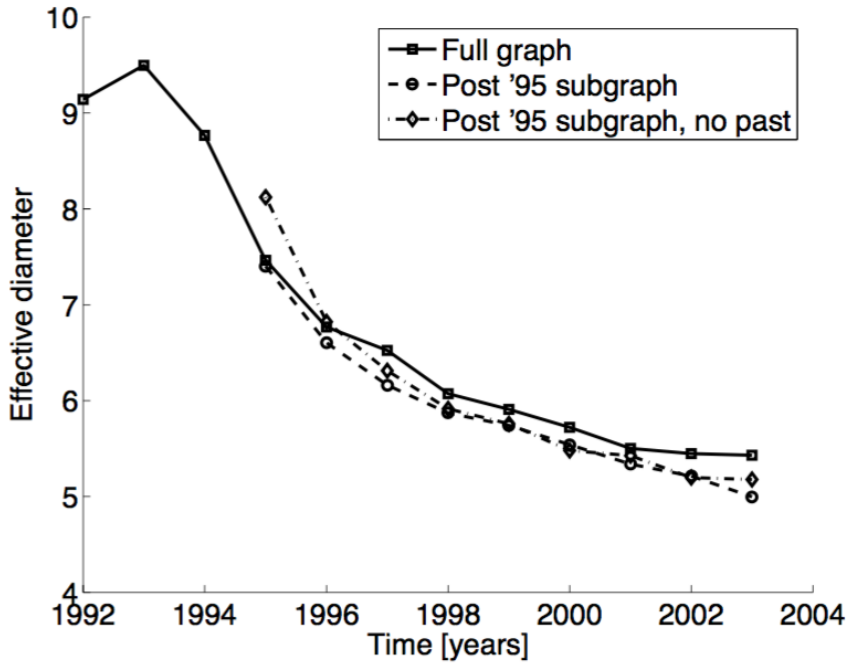
Три варианта графика:

- Весь граф.
- Граф после момента t_0.
- Граф после момента t_0 без «прошлого».

- Для определения эффективного диаметра в силу сложности вычислений на больших графах использовались 2 различных реализации Approximate Neighborhood Function (ANF).

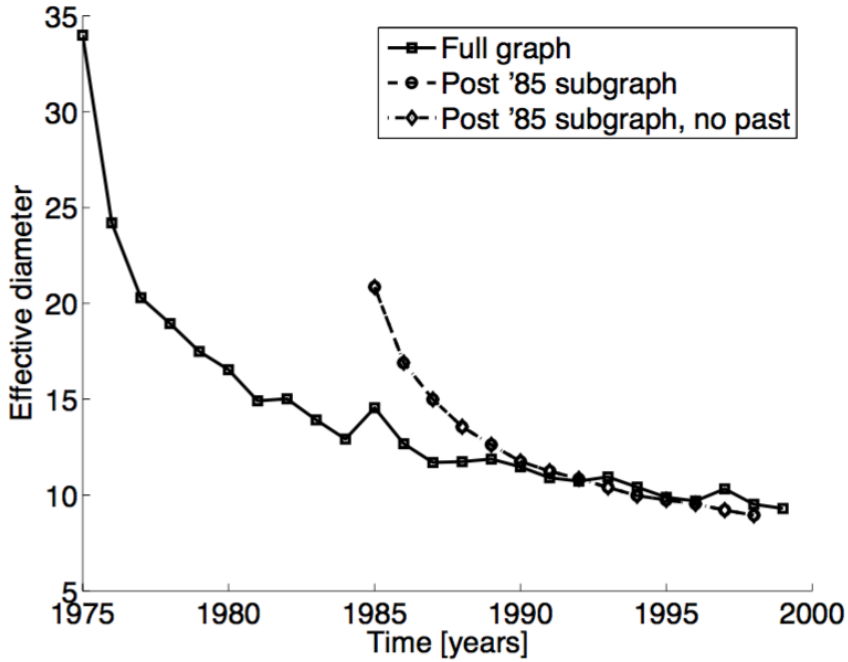
Реальный граф: цитирования в физике

- Цитирование среди статей по физике.
- Данные из arXiv.
- В 1992 году 1239 статей и 2717 цитирований.
- В 2003 году 29555 статей и 352807 цитирований.
- Каждая точка соответствует ситуации на определённый месяц.



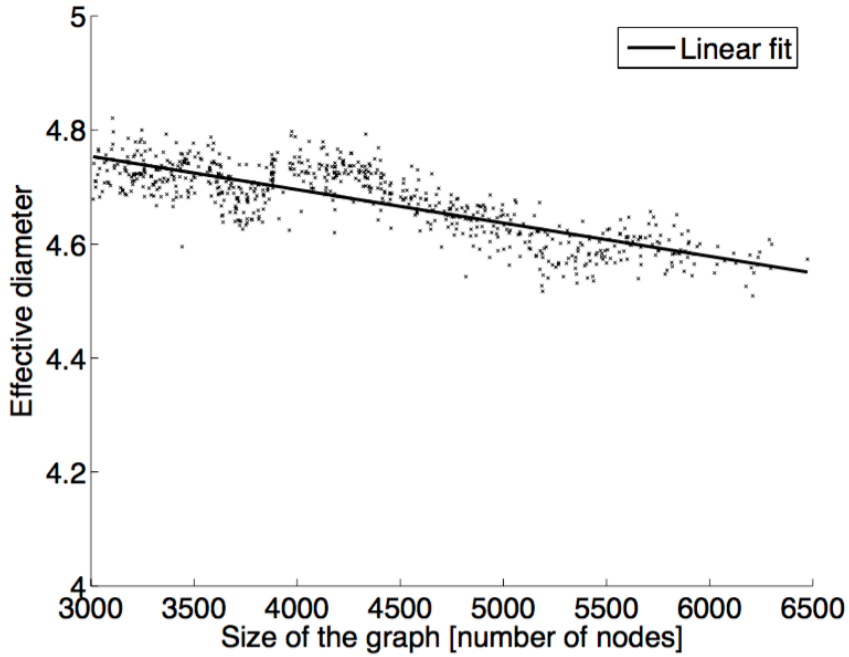
Реальный граф: цитирование в патентах

- Цитирование среди патентов в США.
- В 1975 году 334000 патентов и 676000 цитирований.
- В 1999 году 2.9 миллионов патентов и 16.5 миллионов цитирований.
- Каждая точка соответствует ситуации на определённый год.



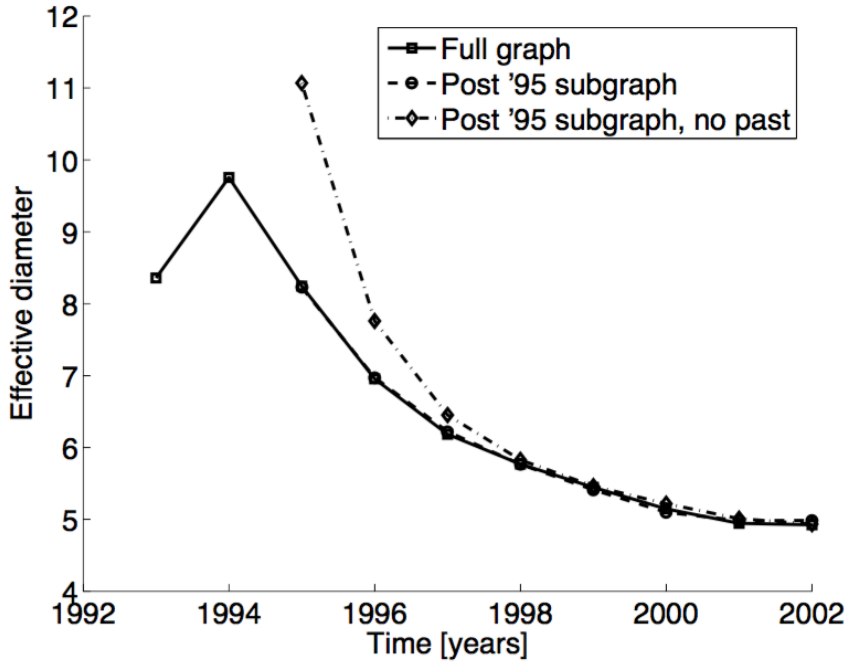
Реальный граф: граф интернета

- Обмен пакетами среди маршрутизаторов.
- Не только добавление вершин и рёбер, но и удаление.
- В 1997 году 3000 вершин и 10000 рёбер.
- В 2000 году 6000 вершин и 26000 рёбер.
- Каждая точка соответствует ситуации на каждый день.



Реальный граф: сооавторство

- Связь между авторами и статьями, в написании которых они участвовали.
- Данные из arXiv по нескольким категориям.
- В 1992 году 318 вершин и 272 ребра.
- В 2000 году 58000 вершин (20000 авторов и 38000 статей) и 133000 рёбер.



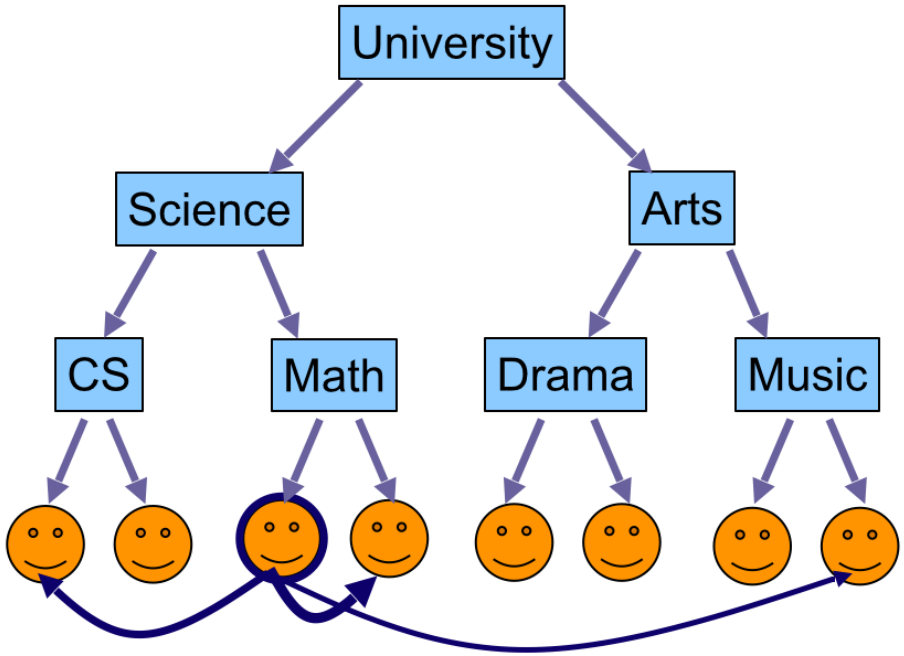
Новые методы генерации социального графа

- Существующие модели генерации графа не удовлетворяют законам увеличения средней степени вершин и уменьшению эффективного диаметра.
- Авторы предложили 2 своих модели для генерации социального графа.

Модель Community Guided Attachment: идея

Представим структуру сообщества.

- Внутригруппых связей много.
- Межгрупповых связей мало.



Модель Community Guided Attachment: построение

- Авторы строят модель, в которой показатель уплотнения α зависит от процесса генерации вершин и рёбер.
- Пусть при появлении новой вершины из неё генерируется $n(t)^{\alpha-1}$ рёбер. Тогда будет выполняться закон увеличения средней степени вершин.
- Авторы берут сбалансированное (b - количество потомков у каждой нелистовой вершины) дерево высоты H. Листья дерева будут узлами социального графа ($n = b^H$).
- Пусть $h(v, w)$ - расстояние между двумя листьями дерева, которое определяется как высота минимального поддерева начального дерева, содержащая листья v и w.

Модель Community Guided Attachment: построение (2)

- Построим на листьях дерева случайных граф, где вероятностью связи между вершинами будет функция $f(h)$ – функция сложности.
- Функция $f(h)$ должна быть убывающей. Авторы предлагают определить её следующим образом:

$$f(h) = c^{-h}$$

- $c \geq 1$ – коэффициент сложности, h - расстояние между листьями в дереве.
- Получим социальный граф, для которого выполняется закон уменьшение средней степени вершин.

Модель Community Guided Attachment

Теорема 1

В графе, построенном по модели Community Guided Attachment, средняя степень вершин пропорциональна:

$$d \propto \begin{cases} n^{1-\log_b c}, & 1 \leq c < b \\ \log_b n, & c = b \\ const, & c > b \end{cases}$$

- Если $1 \leq \ < b$, то $e(t) \propto n(t)^{2-\log_b c}$.
- $c = 1 \Rightarrow a = 2$ – много связей между группами.
- $c = b \Rightarrow a = 1$ – мало связей между группами, линейный рост средней степени вершин.

Модель Dynamic Community Guided Attachment

- Пусть теперь граф будет ориентированным.
- Будем использовать не только листья начального дерева, но и все внутренние вершины.
- Начинаем с дерева из одной вершины.
- В момент времени t из графа высотой $t-1$ делаем граф высотой t, добавлением b листьев к каждому старому листу. Все новые листья становятся новыми вершинами социального графа.
- $d(v,w)$ – растояние от v до ближайшего общего с w предка и от этого предка до w. Если v и w – листья, то $d(v,w) = 2h(v,w)$, где $h(v,w)$ из предыдущего варианта построения.
- Для каждой пары вершин v и w строится ребро с вероятностью $c^{-\frac{d(v,w)}{2}}$.

Зачем ещё одна модель?

- Модель Community Guided Attachment удовлетворяет закону роста средней степени вершин.
- Но хотелось бы ещё, чтобы уменьшался эффективный диаметр.

Мысли для новой модели

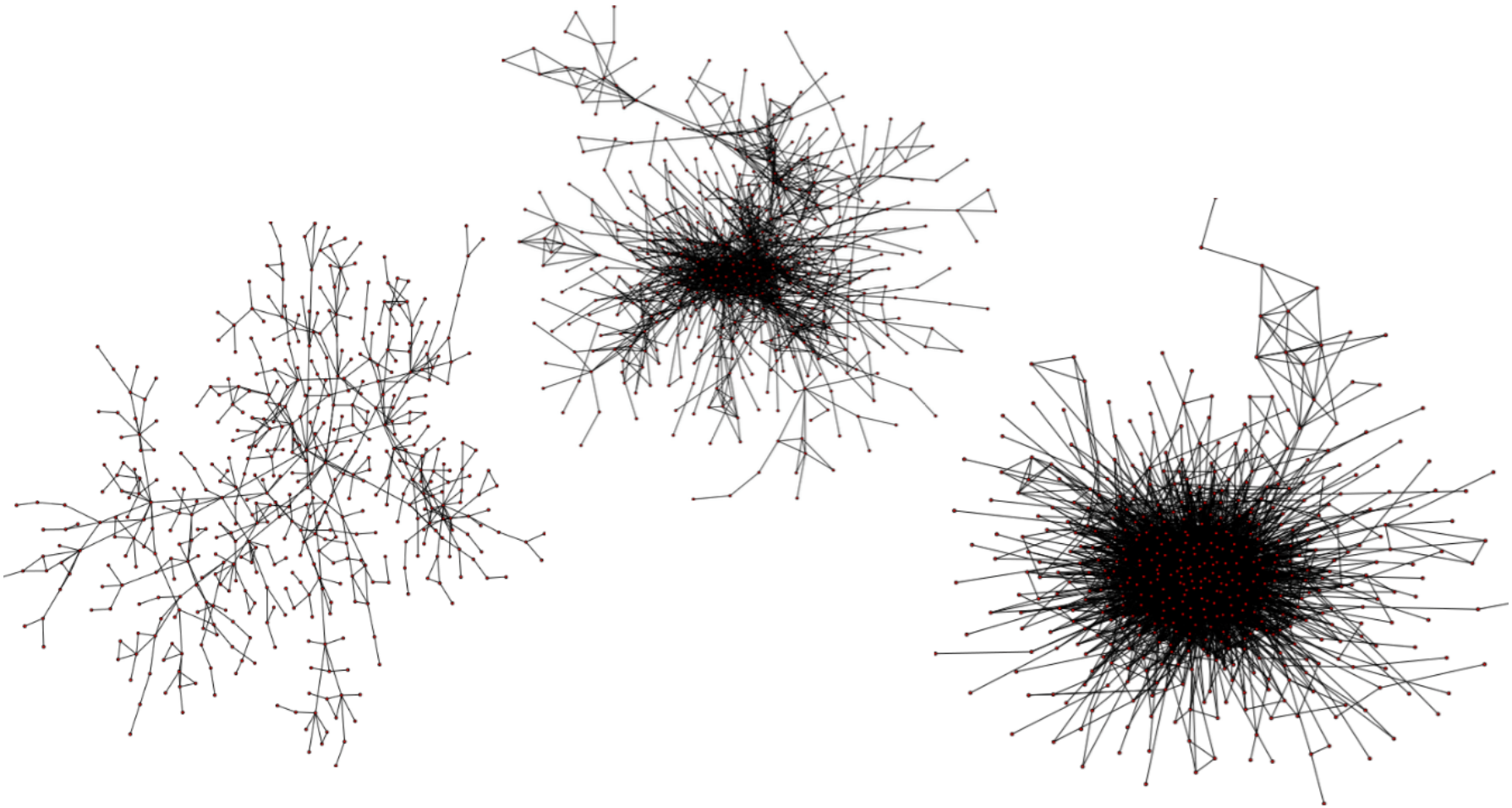
- Как люди добавляют друг друга в друзья? (По мнению авторов статьи.)
  1. Находят одного человека и добавляют его.
  2. Добавляются к некоторым друзьям этого человека.
  3. Повторяют предыдущие пункты рекурсивно.
  4. Иногда просят своих друзей представить себя другим.

- Попробуем скомбинировать:
  1. Процесс «богатые богатеют» для тяжёлых хвостов распределения входящих рёбер.
  2. Модель «копирования» построения графов для образования групп (при этом не использовать явные группы модели Community Guided Attachment).
  3. Модель Community Guided Attachment для выполнения закона увеличения средней степени вершин.
  4. Что-нибудь ещё для уменьшение эффективного диаметра.

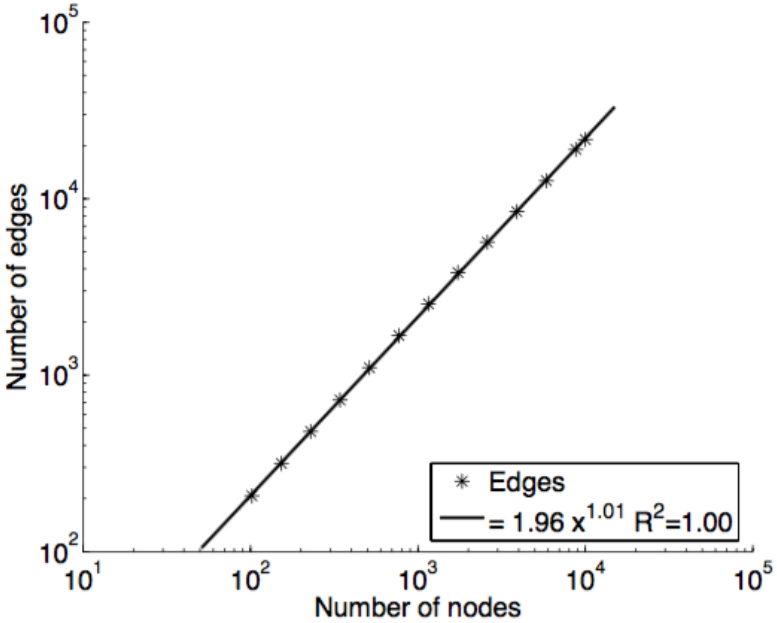
Модель Forest Fire

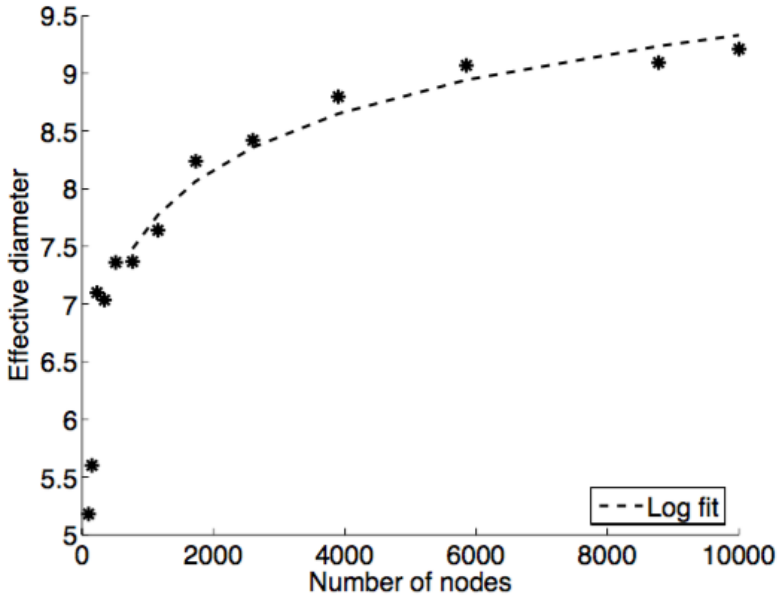
- Необходимо задать 2 параметра: p – вероятность «углубления» (forward burning probability) и r – коэффициент возврата (backward burning ratio).
- Начнём построение с графа из одной вершины.
- Пусть на момент $t > 1$ добавляется вершина v.
- Для вершины v случайно выбирается вершина w (ambassador node) и строится ребро к w.
- Выбирается случайное число x из биномиального расределения с математическим ожидаем $(1-p)^{-1}$. Выбирается x вершин, которые связаны с w так, чтобы входящих в w рёбер из них было в r раз меньше исходящих из w рёбер. Строится ребро из v к этим рёбрам.
- Продолжаем рекурсивно для новых вершин. Так как не заходим повторно в одну и ту же вершину, то процесс не зациклится.

Примеры



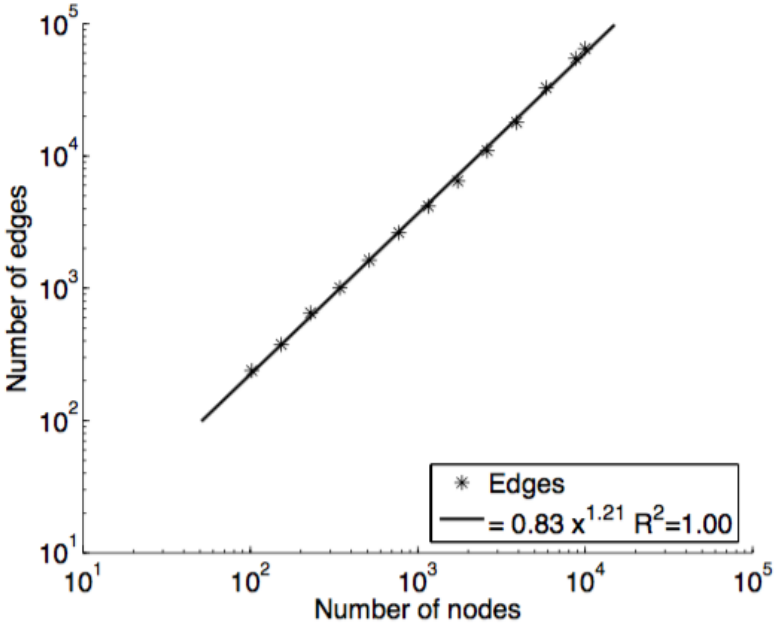
Эксперименты с моделью Forest Fire (1)

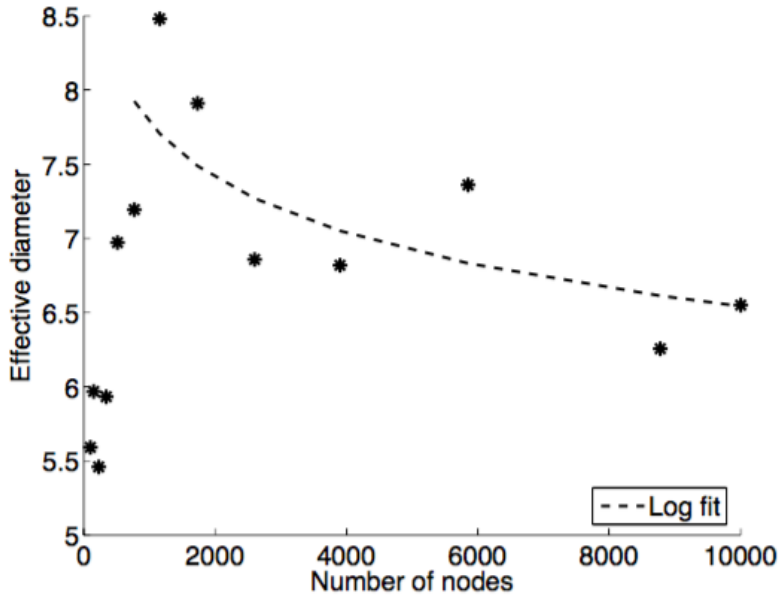




- $\alpha = 1.01$.
- $p = 0.35$.
- $r = 5$.

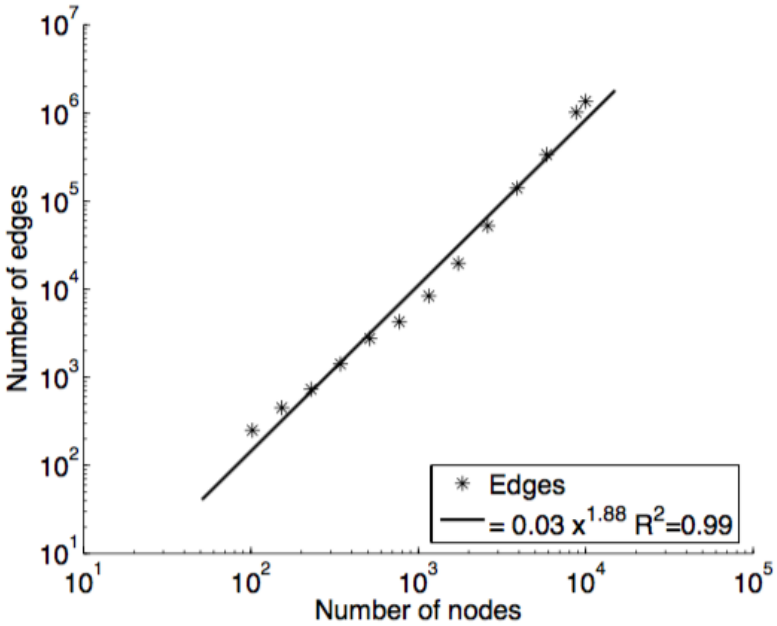
Эксперименты с моделью Forest Fire (2)

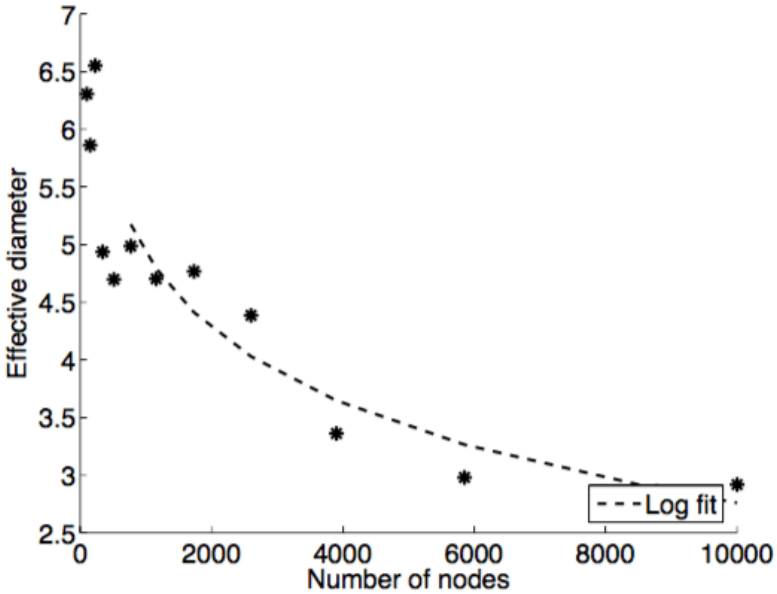




- Наиболее реалистичный случай.
- $\alpha = 1.21$.
- $p = 0.37$.
- $r \approx 3.1$.

Эксперименты с моделью Forest Fire (3)

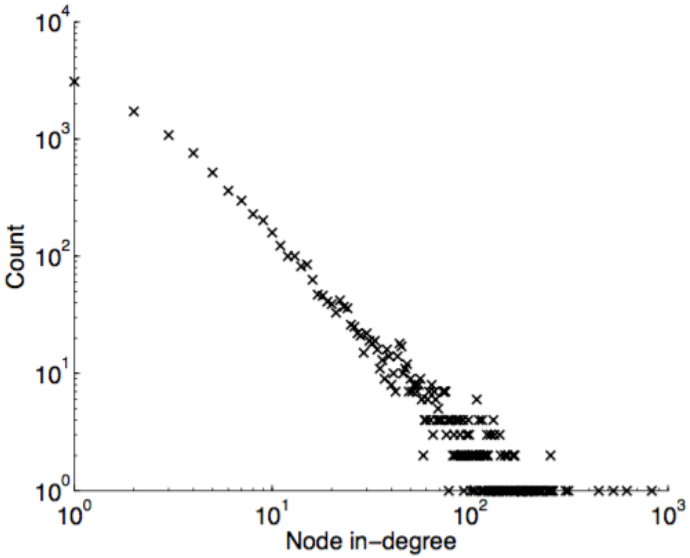




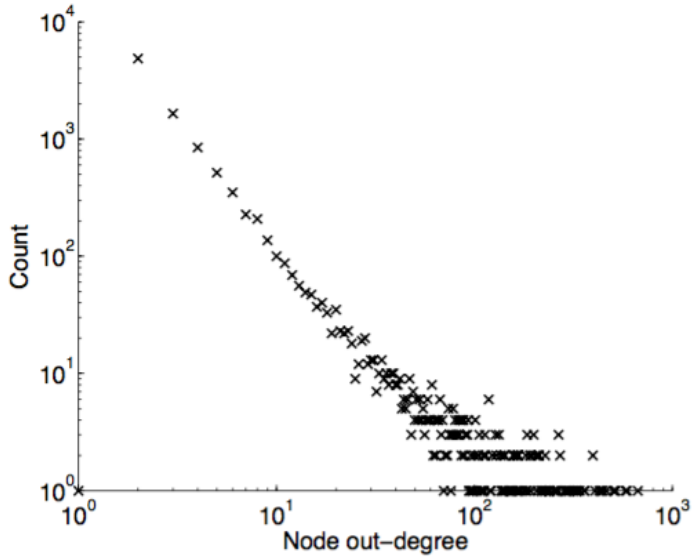
- $\alpha \approx 2.$
- $p = 0.38.$
- $r \approx 2.9.$

Модель Forest Fire

- Требования выполнились (почти).
- Тяжёлые хвости распределений степеней.



In-degree



Out-degree

- Неизвестно, почему уменьшается диаметр.
- Возможные улучшения: вершины-сироты и несколько ambassador nodes.

Выводы

- Закон увеличения средней степени вершин.
- Уменьшение эффективного диаметра.
- Модель Community Guided Attachment с ростом средней степени вершин.
- Модель Forest Fire с ростом средней степени вершин, уменьшением эффективного диаметра и тяжёлыми хвостами распределния степеней вершин.

Список литературы (1)

[1] J. Abello, A. L. Buchsbaum, and J. Westbrook. A functional approach to external graph algorithms. In *Proceedings of the 6th Annual European Symposium on Algorithms*, pages 332–343. Springer-Verlag, 1998.

[2] J. Abello, P. M. Pardalos, and M. G. C. Resende. *Handbook of massive data sets*. Kluwer, 2002.

[3] R. Albert and A.-L. Barabasi. Emergence of scaling in random networks. *Science*, pages 509–512, 1999.

[4] R. Albert, H. Jeong, and A.-L. Barabasi. Diameter of the world-wide web. *Nature*, 401:130–131, September 1999.

[5] Z. Bi, C. Faloutsos, and F. Korn. The dgx distribution for mining massive, skewed data. In *KDD*, pages 17–26, 2001.

[6] B. Bollobas and O. Riordan. The diameter of a scale-free random graph. *Combinatorica*, 24(1):5–34, 2004.

[7] A. Broder, R. Kumar, F. Maghoul, P. Raghavan, S. Rajagopalan, R. Stata, A. Tomkins, and J. Wiener. Graph structure in the web: experiments and models. In *Proceedings of World Wide Web Conference*, 2000.

[8] D. Chakrabarti, Y. Zhan, and C. Faloutsos. R-mat: A recursive model for graph mining. In *SDM*, 2004.

[9] F. Chung and L. Lu. The average distances in random graphs with given expected degrees. *Proceedings of the National Academy of Sciences*, 99(25):15879–15882, 2002.

[10] C. Cooper and A. Frieze. A general model of web graphs. *Random Struct. Algorithms*, 22(3):311–335, 2003.

[11] M. Faloutsos, P. Faloutsos, and C. Faloutsos. On power-law relationships of the internet topology. In *SIGCOMM*, pages 251–262, 1999.

[12] J. Gehrke, P. Ginsparg, and J. M. Kleinberg. Overview of the 2003 kdd cup. *SIGKDD Explorations*,

[13] B. H. Hall, A. B. Jaffe, and M. Trajtenberg. The nber patent citation data file: Lessons, insights and methodological tools. NBER Working Papers 8498, National Bureau of Economic Research, Inc, Oct. 2001.

[14] B. A. Huberman and L. A. Adamic. Growth dynamics of the world-wide web. *Nature*, 399:131, 1999.

[15] J. S. Katz. The self-similar science system. *Research Policy*, 28:501–517, 1999.

[16] J. S. Katz. Scale independent bibliometric indicators. *Measurement: Interdisciplinary Research and Perspectives*, 3:24–28, 2005.

[17] J. M. Kleinberg. Small-world phenomena and the dynamics of information. In *Advances in Neural Information Processing Systems 14*, 2002.

[18] J. M. Kleinberg, R. Kumar, P. Raghavan, S. Rajagopalan, and A. Tomkins. The web as a graph: Measurements, models, and methods. In *Proc. International Conference on Combinatorics and Computing*, pages 1–17, 1999.

[19] R. Kumar, P. Raghavan, S. Rajagopalan, D. Sivakumar, A. Tomkins, and E. Upfal. Stochastic models for the web graph. In *Proc. 41st IEEE Symp. on Foundations of Computer Science*, 2000.

[20] R. Kumar, P. Raghavan, S. Rajagopalan, and A. Tomkins. Trawling the web for emerging cyber-communities. In *Proceedings of 8th International World Wide Web Conference*, 1999.

[21] F. Menczer. Growing and navigating the small world web by local content. *Proceedings of the National Academy of Sciences*, 99(22):14014–14019, 2002.

[22] S. Milgram. The small-world problem. *Psychology Today*, 2:60–67, 1967.

[23] M. Mitzenmacher. A brief history of generative models for power law and lognormal distributions, 2004.

Список литературы (2)

[24] M. E. J. Newman. The structure and function of complex networks. *SIAM Review*, 45:167–256, 2003.

[25] A. Ntoulas, J. Cho, and C. Olston. What's new on the web? the evolution of the web from a search engine perspective. In *WWW Conference*, pages 1–12, New York, New York, May 2004.

[26] U. of Oregon Route Views Project. Online data and reports. http://www.routeviews.org.

[27] C. R. Palmer, P. B. Gibbons, and C. Faloutsos. Anf: A fast and scalable tool for data mining in massive graphs. In *SIGKDD*, Edmonton, AB, Canada, 2002.

[28] S. Redner. Citation statistics from more than a century of physical review. Technical Report physics/0407137, arXiv, 2004.

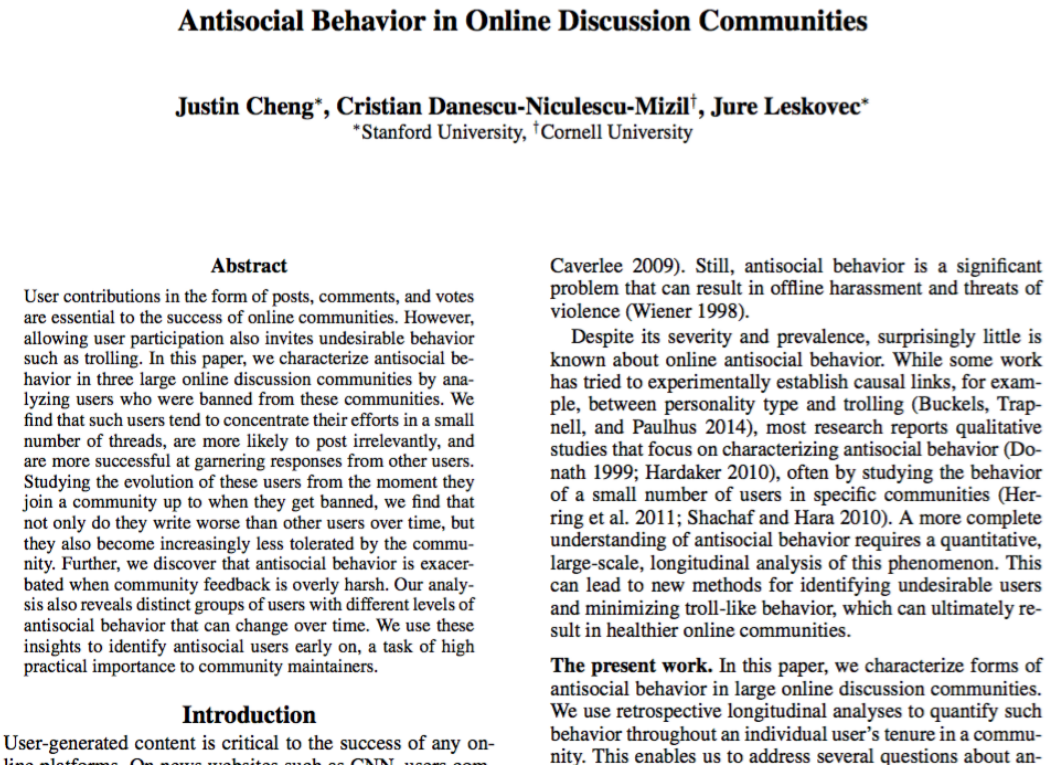
[29] M. Schroeder. *Fractals, Chaos, Power Laws: Minutes from an Infinite Paradise*. W.H. Freeman and Company, New York, 1991.

[30] D. J. Watts, P. S. Dodds, and M. E. J. Newman. Collective dynamics of 'small-world' networks. *Nature*, 393:440–442, 1998.

[31] D. J. Watts, P. S. Dodds, and M. E. J. Newman. Identity and search in social networks. *Science*, 296:1302–1305, 2002.

Содержание

Antisocial Behavior in Online Discussion Communities



Antisocial Behavior in Online Discussion Communities

Justin Cheng*, **Cristian Danescu-Niculescu-Mizil**†, **Jure Leskovec***
*Stanford University, †Cornell University

Abstract

User contributions in the form of posts, comments, and votes are essential to the success of online communities. However, allowing user participation also invites undesirable behavior such as trolling. In this paper, we characterize antisocial behavior in three large online discussion communities by analyzing users who were banned from these communities. We find that such users tend to concentrate their efforts in a small number of threads, are more likely to post irrelevantly, and are more successful at garnering responses from other users. Studying the evolution of these users from the moment they join a community up to when they get banned, we find that not only do they write worse than other users over time, but they also become increasingly less tolerated by the community. Further, we discover that antisocial behavior is exacerbated when community feedback is overly harsh. Our analysis also reveals distinct groups of users with different levels of antisocial behavior that can change over time. We use these insights to identify antisocial users early on, a task of high practical importance to community maintainers.

Introduction

User-generated content is critical to the success of any on-

Caverlee 2009). Still, antisocial behavior is a significant problem that can result in offline harassment and threats of violence (Wiener 1998).

Despite its severity and prevalence, surprisingly little is known about online antisocial behavior. While some work has tried to experimentally establish causal links, for example, between personality type and trolling (Buckels, Trapnell, and Paulhus 2014), most research reports qualitative studies that focus on characterizing antisocial behavior (Donath 1999; Hardaker 2010), often by studying the behavior of a small number of users in specific communities (Herring et al. 2011; Shachaf and Hara 2010). A more complete understanding of antisocial behavior requires a quantitative, large-scale, longitudinal analysis of this phenomenon. This can lead to new methods for identifying undesirable users and minimizing troll-like behavior, which can ultimately result in healthier online communities.

The present work. In this paper, we characterize forms of antisocial behavior in large online discussion communities. We use retrospective longitudinal analyses to quantify such behavior throughout an individual user's tenure in a community. This enables us to address several questions about an-

- Jure Leskovec, Justin Cheng, Cristian Danescu-Niculescu-Mizil.
- AAAI International Conference on Weblogs and Social Media (ICWSM), 2015. Best paper award honorable mention.

Введение

- Онлайн-сообщества поддерживаются за счёт постов, комментариев и голосов.
- Некоторые люди начинают вести себя нежелательным образом, нарушая органичность сообщества.
- Сообщество жалуется или «минусует» этих людей. Модератор проверяет жалобы и банит пользователей.
- Что насчёт автоматизации?

Вопросы

- Люди со временем приобретают антисоциальное поведение в сообществе или оно является врождённым?
- Реакция сообщества помогает избавиться от антисоциальных пользователей или все люди становятся более антисоциальными?
- Можно ли определить антисоциальных пользователей заранее?

Какие данные будем использовать?

- Источники:
  1. CNN.com – новостной сайт.
  2. Breitbart.com – политический новостной сайт.
  3. IGN.com – сайт про компьютерные игры.
- За более 18 месяцев 1.7 миллионов пользователей написали около 40 миллионов постов и оставили более 100 миллионов голосов.
- Постом называется комментарий или ответ к комментарию пользователя к статье.
- Пользователи, систематически нарушающие правила, блокируются навсегда.

Community	# Users	# Users Banned	# Threads	# Posts	# Posts Deleted	# Posts Reported
CNN	1,158,947	37,627 (3.3%)	200,576	26,552,104	533,847 (2.0%)	1,146,897 (4.3%)
IGN	343,926	5,706 (1.7%)	682,870	7,967,414	184,555 (2.3%)	88,582 (1.1%)
Breitbart	246,422	5,350 (2.2%)	376,526	4,376,369	119,265 (2.7%)	117,779 (2.7%)

Первая попытка

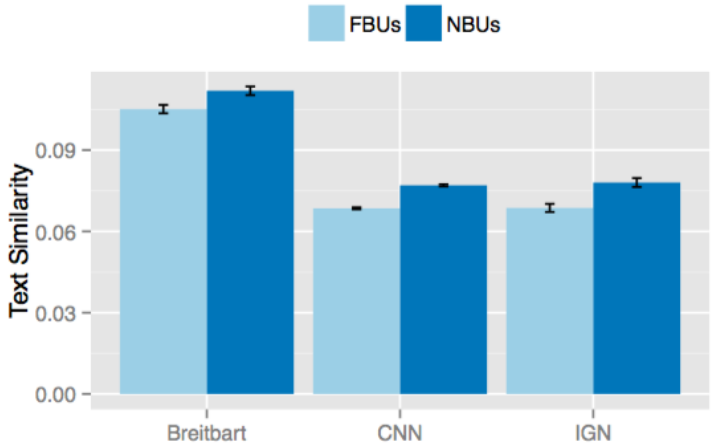
- Возьмём примеры сообщений пользователей двух групп (FBUs – Future-Banned Users, NBUs – Never-Banned Users) в каждом источнике по 500 штук.
- Ручным трудом разметим эти данные: каждое сообщение по пятибалльной шкале оценит по три человека на наличие оскорблений, ненормативной лексики, спама.
- Получили, что у FBUs средний рейтинг сообщений ниже, чем у NBUs (2.4 против 3.0).
- На размеченных данных по биграммам обучим логистическую регрессию. Посты с рейтином выше 3.0 будем считать нормальными, прочие относить к негативным.
- AUC = 0.7.

Схожесть постов с предыдущими постами

- Сравним схожесть слов в посте с тремя предыдущими постами в той же теме.
- Используем косинусный коэффициент (коэффициент Отиаи):

$$K = \frac{|A \cap B|}{\sqrt{|A||B|}}$$

- Но корреляция между схожестью слов в постах и удалением постов слишком слабая (отрицательная).



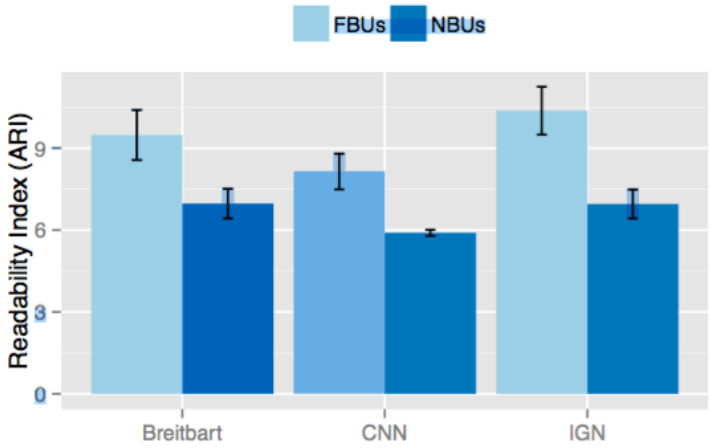
(a) Text Sim. of a Post with its Parent Thread

Коэффициент понятности текста

- Сравним посты обеих групп пользователь по понятности.
- Используем Automated readability index:

$$ARI = \frac{\#words}{\#sentences} + 9\frac{\#charachers}{\#words}$$

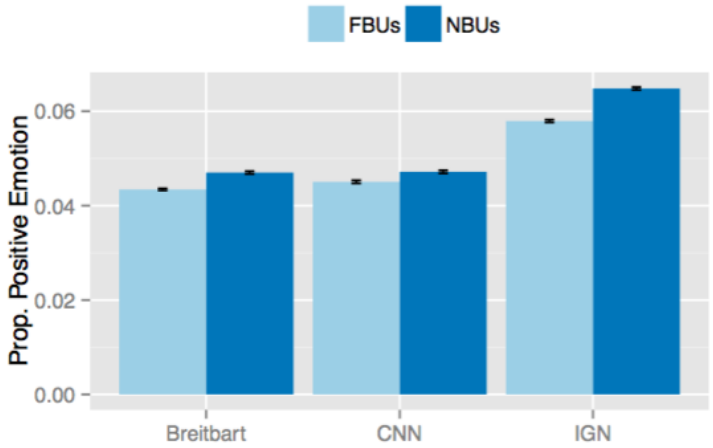
- Вывод: посты FBUs менее понятны.



(b) Readability Index

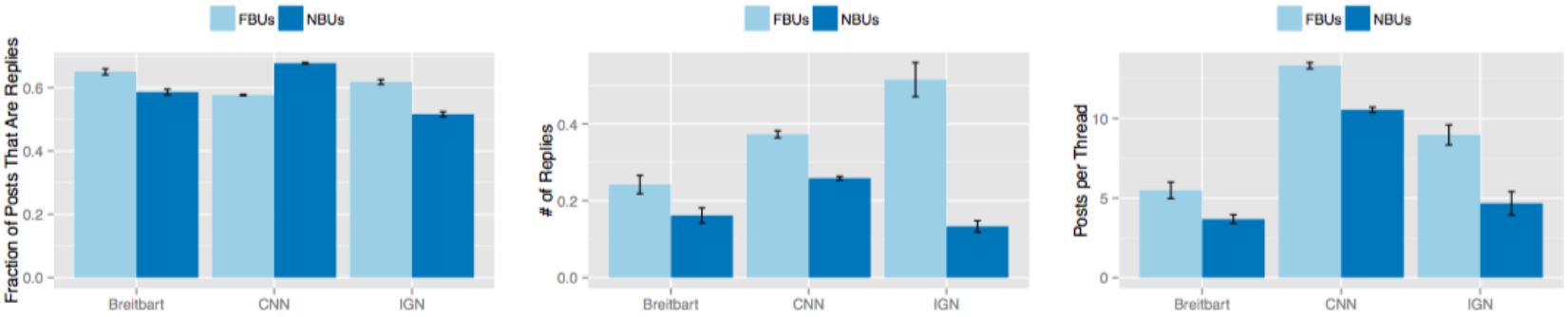
Позитивные эмоции

- Сравним посты обеих групп пользователь по позитивности эмоций.
- Используем LIWC – библиотека слов по эмоциям и прочим категориям.
- Вывод: посты FBUs менее позитивны.



(c) Positive Emotion

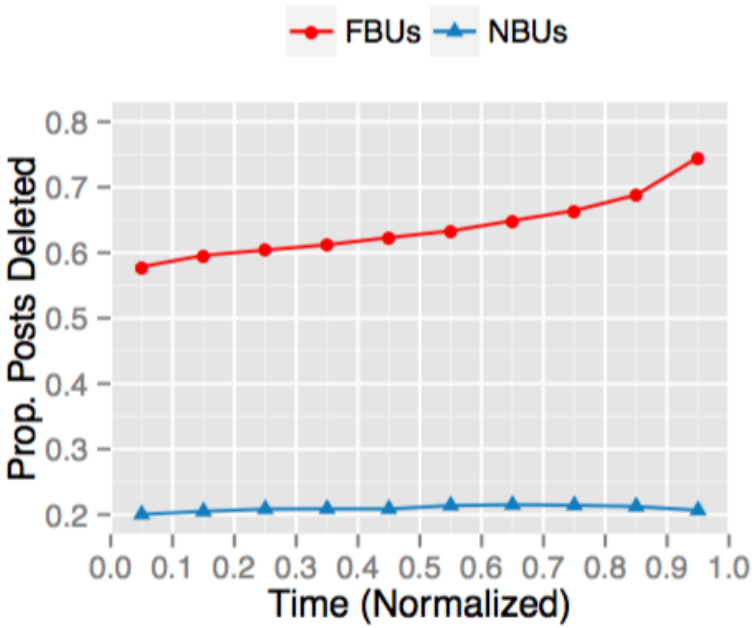
Активность



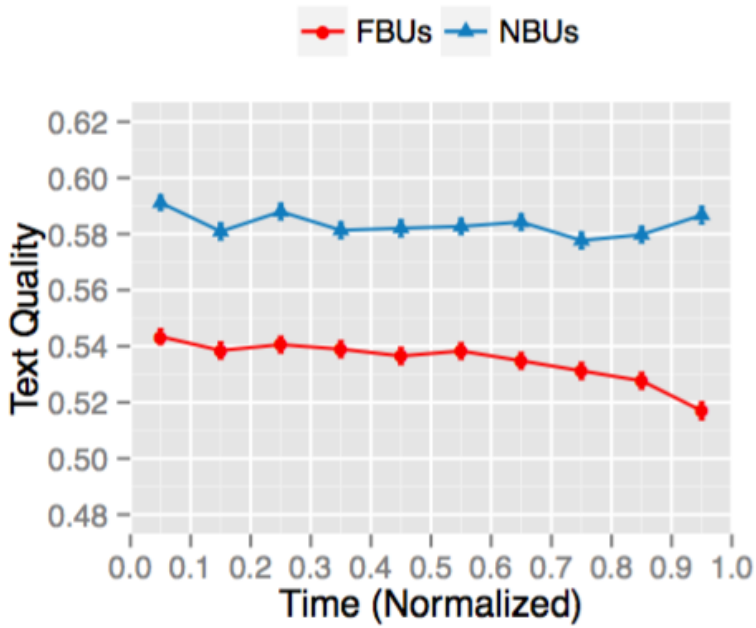
(a) Fraction of Posts That Are Replies (b) # Replies (c) Posts per Thread

- В Breitbart и IGN FBUs чаще отвечают другим пользователям, а на CNN предпочитают начинать новую дискуссию чаще, чем остальные пользователи.
- FBUs больше отвечают на соощения других пользователей, чем другие пользователи.
- FBUs оставляют больше постов внутри одной темы, чем другие пользователи.

Изменение во времени



(a) Post deletion rate



(b) Text quality

Но что происходит на самом деле?

- FBUs пишут со временем хуже и хуже.
- Сообщетсво со временем начинает узнавать таких пользователей и меньше терпит их поведение.

Первая гипотеза

- Действительно ли FBUs начинают писать хуже и хуже?
- Снова используем ручной труд для проверки.

	Mean Post Appropriateness on CNN (1-5)		
	All Posts	**First 10%**	**Last 10%**
FBUs	2.7	3.0	2.3
NBUs	3.3	3.5	3.2

- Вывод: качество постов ухудшается, но у всех пользователей. Для FBUs изменение заметнее.

Вторая гипотеза

- Действительно ли сообщество со временем меньше и меньше терпит посты FBUs?
- Составим пары постов для одного пользователя одного качества, где одно выбрано из 10% первых постов, а второе из последних 10%.
- Используем Критерий Уилкоксона для связных выборок. Получаем, что среди последних 10% постов у FBUs больше шансов на удаление, чем у NBUs, несмотря на сохранение качества постов.
- Кроме того, замечено, что если среди двух пользователей, начавших писать посты одного качества, у одного были удалённые, то со временем у него качество постов будет выше, чем у второго.

Признаки

Feature Set	Features
Post (20)	number of words, readability metrics (e.g., ARI), LIWC features (e.g., affective)
Activity (6)	posts per day, posts per thread, largest number of posts in one thread, fraction of posts that are replies, votes given to other users per post written, proportion of up-votes given to other users
Community (4)	votes received per post, fraction of up-votes receieved, fraction of posts reported, number of replies per post
Moderator (5)	fraction of posts deleted, slope and intercept of linear regression lines (i.e., m_1,m_2,c_1,c_2)

Методика

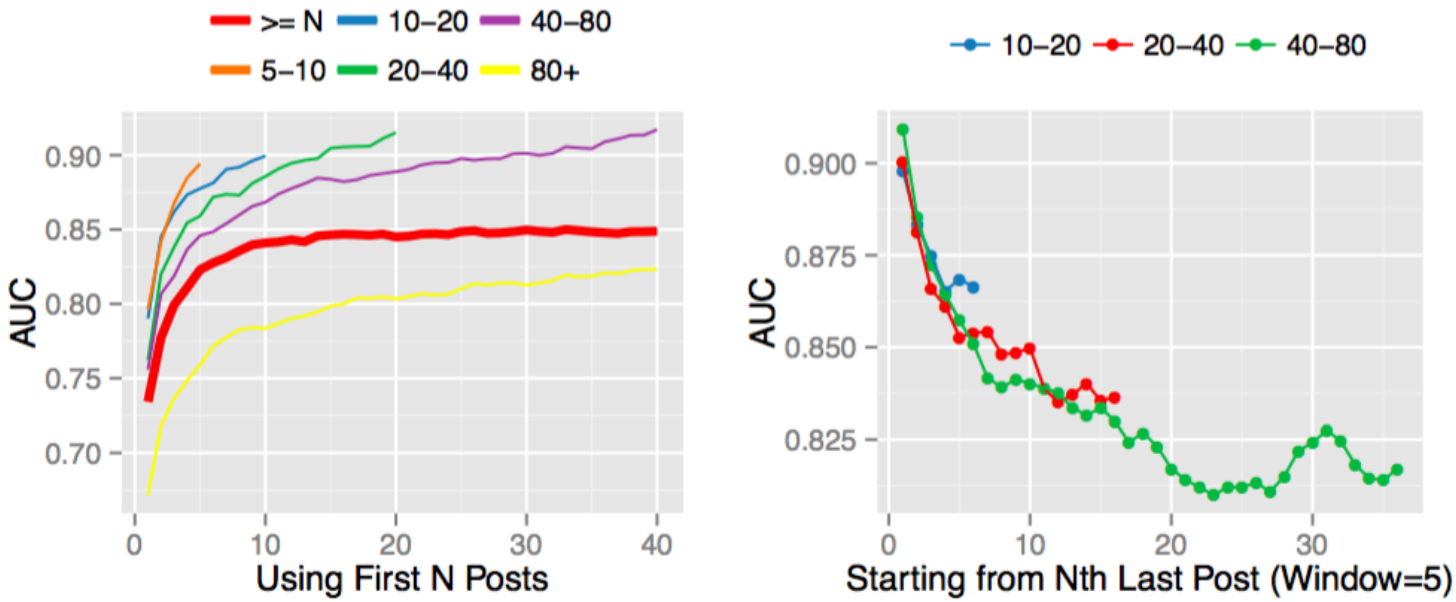
- 10 первых постов пользователя в сбалансированной по количеству FBUs и NBUs выборке.
- Random forest.
- 10-fold cross validation.

Результаты

	CNN	IGN	Breitbart
Bag-of-words	0.70	0.72	0.65
Prop. Deleted Posts	0.74	0.72	0.72
Post	0.62	0.67	0.58
+ Activity	0.73 (0.66)	0.74 (0.65)	0.66 (0.64)
+ Community	0.83 (0.75)	0.79 (0.72)	0.75 (0.69)
+ Moderator	0.84 (0.75)	0.83 (0.73)	0.78 (0.72)

Что-нибудь ещё?

- Если смотреть только на 5 первых постов, то AUC = 0.8 (было 0.82).
- Что если брать большее число постов? Что если брать посты не из первых?



(a) Performance against # posts (b) Performance against time

Обобщение классификатора

Что если применить обученный классификатор к другому сообществу?

		Trained on		
		CNN	**IGN**	**Breitbart**
Tested on	**CNN**	0.84	0.74	0.76
	IGN	0.69	0.83	0.74
	Breitbart	0.74	0.75	0.78

Список литературы

Adler, B. T.; De Alfaro, L.; Mola-Velasco, S. M.; Rosso, P.; and West, A. G. 2011. Wikipedia vandalism detection: Combining natural language, metadata, and reputation features. In *CICLing*.

Baker, P. 2001. Moral panic and alternative identity construction in Usenet. *J Comput-Mediat Comm*.

Binns, A. 2012. DON'T FEED THE TROLLS! Managing troublemakers in magazines' online communities. *Journalism Practice*.

Blackburn, J., and Kwak, H. 2014. STFU Noob!: Predicting crowdsourced decisions on toxic behavior in online games. In *WWW*.

Buckels, E. E.; Trapnell, P. D.; and Paulhus, D. L. 2014. Trolls just want to have fun. *Pers Indiv Differ*.

Cheng, J.; Danescu-Niculescu-Mizil, C.; and Leskovec, J. 2014. How community feedback shapes user behavior. In *ICWSM*.

Chesney, T.; Coyne, I.; Logan, B.; and Madden, N. 2009. Griefing in virtual worlds: causes, casualties and coping strategies. *Inform Syst J*.

Danescu-Niculescu-Mizil, C.; West, R.; Jurafsky, D.; Leskovec, J.; and Potts, C. 2013. No country for old members: User lifecycle and linguistic change in online communities. In *WWW*.

Danescu-Niculescu-Mizil, C.; Gamon, M.; and Dumais, S. 2011. Mark my words! Linguistic style accommodation in social media. In *WWW*.

Diakopoulos, N., and Naaman, M. 2011. Towards quality discourse in online news comments. In *CSCW*.

Donath, J. S. 1999. Identity and deception in the virtual community. *Communities in cyberspace*.

Hardaker, C. 2010. Trolling in asynchronous computer-mediated communication: From user discussions to academic definitions. *J Politeness Res*.

Hardaker, C. 2013. Uh.... not to be nitpicky, but... the past tense of drag is dragged, not drug. *JLAC*.

Herring, S.; Job-Sluder, K.; Scheckler, R.; and Barab, S. 2011. Searching for safety online: Managing "trolling" in a feminist forum. *The Information Society*.

Hsu, C.-F.; Khabiri, E.; and Caverlee, J. 2009. Ranking comments on the social web. In *CSE*.

Javanmardi, S.; McDonald, D. W.; and Lopes, C. V. 2011. Vandalism detection in Wikipedia: a high-performing, feature-rich model and its reduction through Lasso. In *WikiSym*.

Juvonen, J., and Gross, E. F. 2008. Extending the school grounds? Bullying experiences in cyberspace. *J School Health*.

Kirman, B.; Lineham, C.; and Lawson, S. 2012. Exploring mischief and mayhem in social computing or: how we learned to stop worrying and love the trolls. In *CHI EA*.

Laniado, D.; Kaltenbrunner, A.; Castillo, C.; and Morell, M. F. 2012. Emotions and dialogue in a peer-production community: the case of wikipedia. In *WikiSym*.

Pennebaker, J. W.; Francis, M. E.; and Booth, R. J. 2001. Linguistic inquiry and word count: LIWC 2001.

Potthast, M.; Stein, B.; and Gerling, R. 2008. Automatic vandalism detection in Wikipedia. In *Lect Notes Comput Sc*.

Rosenbaum, P. R., and Rubin, D. B. 1983. The central role of the propensity score in observational studies for causal effects. *Biometrika*.

Shachaf, P., and Hara, N. 2010. Beyond vandalism: Wikipedia trolls. *J Inf Sci*.

Sood, S. O.; Churchill, E. F.; and Antin, J. 2012. Automatic identification of personal insults on social news sites. *ASIST*.

Suler, J. R., and Phillips, W. L. 1998. The bad boys of cyberspace: Deviant behavior in a multimedia chat community. *Cyberpsychol Behav*.

Suler, J. 2004. The online disinhibition effect. *Cyberpsychol Behav*.

Wang, W. Y., and McKeown, K. R. 2010. "Got you!": Automatic vandalism detection in Wikipedia with web-based shallow syntactic-semantic modeling. In *COLING*.

Wiener, D. 1998. Negligent publication of statements posted on electronic bulletin boards: Is there any liability left after Zeran? *Santa Clara L. Rev.*