If walls could talk: Patterns and anomalies in Facebook wallposts

Pravallika Devineni *, Danai Koutra †, Michalis Faloutsos *, Christos Faloutsos †
* Department of Computer Science, University of New Mexico, Albuquerque, NM, USA
†School of Computer Science, Carnegie Mellon University, 5000 Forbes Avenue, Pittsburgh, PA, USA
{pdevineni, michalis}@cs.unm.edu {danai, christos}@cs.cmu.edu

***Abstract*—How do people interact with their Facebook wall? At a high level, this question captures the essence of our work. While most prior efforts focus on Twitter, the much fewer Facebook studies focus on the friendship graph or are limited by the amount of users or the duration of the study. In this work, we model Facebook user behavior: we analyze the wall activities of users focusing on identifying common patterns and surprising phenomena. We conduct an extensive study of roughly *7K* users over *three years* during four month intervals each year. We propose *PowerWall*, a lesser known heavy-tailed distribution to fit our data. Our key results can be summarized in the following points. First, we find that many wall activities, including number of posts, number of likes, number of posts of type photo, *etc.*, can be described by the *PowerWall* distribution. What is more surprising is that most of these distributions have similar slope, with a value close to 1! Second, we show how our patterns and metrics can help us spot surprising behaviors and anomalies. For example, we find a user posting every two days, exactly the same count of posts; another user posting at midnight, with no other activity before or after. Our work provides a solid step towards a systematic and quantitative wall-centric profiling of Facebook user activity.**

I. Introduction

"What can a Facebook wall reveal about the user and their activities?" As Facebook is still a dominant social network, its walls capture a significant part of online social interactions. A challenge in analyzing Facebook behavior is that there are many different activities which include user posts, comments, likes, wall-to-wall interactions with friends, Facebook games and applications. Furthermore, each Facebook user has a wall, which we can describe as their "space", and a newsfeed which aggregates information from the walls of the user's friends. Here we focus on the activities as they appear on the wall of a user, and we attempt to detect common, persistent and surprising behaviors.

We focus on modeling the posting behavior of Facebook users on their walls. This topic has received relatively little attention. On one hand, most previous work on Facebook studies the evolution of friendship network [1], [2], [3], the ego network of a user [4], [5], the propagation of benign and malicious cascades [6], [7]. On the other hand, studies on user profiling have typically focused on other social platforms and communication networks, such as Twitter [8], the telephone call network [9], [10], [11], blogs [12] etc. More recently, there have been several studies that extract real-life information from users to predict conditions like depression [13] and stress [14]. We discuss related work in more detail in section VI.

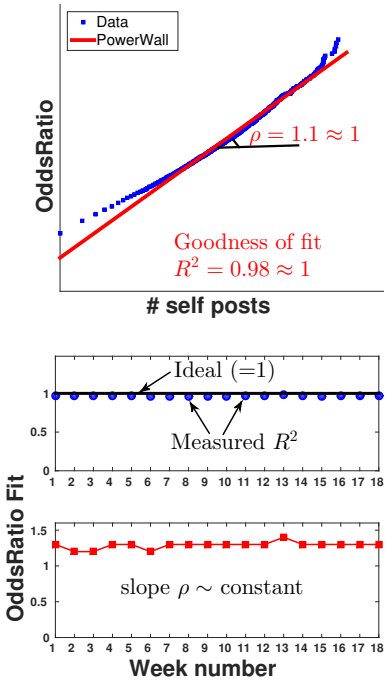


Fig. 1. ***PowerWall* fits real data well, and persistently**: (Top) The Odds Ratio plot of the number of self posts (feature F_2) in dataset *D-13* fitted with *PowerWall* distribution. We observe that the Odds Ratio fits the empirical data very well with $\rho \approx 1$. The fitting is represented by R^2 and the slope ρ. (Bottom) The plot maintains roughly the same slope and a good approximations fit for each week of the four month period.

In this paper, we study the user interaction and activities on Facebook walls empirically to identify common behaviors, patterns and surprising phenomena. We conduct an extensive study of roughly *7K distinct users* in *four* month intervals for *three different* years.

Our key results are the following:

- **Distribution - *PowerWall***: Almost all the distributions of user activities, such as count of posts, "likes" and comments on users' Facebook walls, obey our *PowerWall* distribution. Figure 1(a) shows the Odds Ratio distribution of the count of self posts created by the user on their own

wall in a four month duration in 2013. The distribution is well approximated by a straight-line in this log-log plot ($R^2 > 0.98$). This approximation holds true for nine different behavioral features, as we discuss in section IV.

- **Pervasiveness of *PowerWall***: Despite the diversity, almost all the behavioral features have similar slope in their odds-ratio plot, in the range of $0.9-1.2$; moreover, this slope is rather stable over time. Figure 1 (b) shows the persistence of slope of the *PowerWall* over three datasets spanning three years, and a measure of the goodness of the fit R^2, which is close to 1 consistently, as we discuss in section IV.
- **Usefulness**: The pervasiveness of our *PowerWall* gives us confidence to spot outliers, like a user posting every two days, exactly the same count of posts; and another user posting at midnight, with no other activity before or after as we discuss in section V.

Reproducibility: We have open-sourced our code, at *https://github.com/pdevineni/Powerwall*. We will also provide controlled access to our data, respecting the privacy of the users.

II. Models and Background

In this section, we provide background on modeling real user behaviors and activities using skewed distributions with heavy tails.

A. Power laws

Power law distributions occur often, including settings in computer science and social sciences. Mathematically, the distribution of a quantity x obeys a power law, if its probability distribution function is of the form

$$p(x) \propto x^{-\alpha},$$

where α is a constant (the exponent of the law). Power-law degree distributions characterize the inter-domain Internet topology with $\alpha = 2.1 - 2.2$ [15]. Power laws have been used to model graph evolution using a process of preferential attachment, a mechanism where newly arriving nodes have a tendency to connect with already well-connected nodes with α in the range $2.1 - 4$ [16], citation networks with $\alpha = 3$ [17], call graphs [9], [11], online social networks like Flicker, Orkut and YouTube and many more. Clauset et al. [18] provide a more expansive list of work on power law distributions.

B. Beyond power laws

Where traditional power laws fall short in explaining deviating behaviors, researchers used heavy-tailed Pareto-like distributions such as lognormal, double Pareto, Weibull distributions to model complex networks and natural phenomena. The log-normal distribution is used for modeling growth in any process where the underlying change rate is a random factor over a time step independent of the current size [19]. The double Pareto lognormal (DPLN) distribution is generated by mixing a number of lognormals and exhibits power law behavior at both the upper and the lower tails. An earlier work [20] used DPLN to model mobile call graphs and identify outlier behaviors. In [10], the authors propose the TLAC distribution, a truncated version of log-logistic distribution to model call duration of users whose distribution comprises of semi-heavy tails. A more recent work [21] proposed Self-Feeding Process that behaves like Poisson in the short term and gives power-law tails in the long term to model dependencies between inter-arrival times in social networks. Weibull distribution is another heavy tailed distribution widely used in reliability engineering to assess the life cycle of products [22]. However, due to the difficulty involved in estimating parameters, it is often left resort from a modeling point of view.

C. The log-logistic distribution

The log-logistic distribution is used in economics and was first introduced in 1952 [23]. The log-logistic distribution presents a modified version of the well-known phenomenon "rich get richer". Table I gives the major symbols we use and their definitions.

TABLE I. Symbols and Definitions

Symbol	Definition
$f_X(.)$	Probability density function (PDF) of a random variable X
$f_D(.)$	PDF of a distribution D e.g. $f_{LL}(x)$ = PDF of log-logistic
$F_X(.)$	Cumulative distribution function (CDF) of a random variable X
$F_D(.)$	CDF of a distribution D e.g. $F_{LL}(x)$ = CDF of log-logistic
$OR_D(.)$	Odds ratio function of a distribution
σ	shape parameter of our distribution
μ	location parameter of our distribution
ρ	slope of the tail of our distribution
α	scaling exponent of the power-law
R^2	coefficient of determination indicates how well the data fits the statistical model
PDF	Probability density function
CDF	Cumulative density function
CCDF	Complementary cumulative density function
OR	Odds ratio function of a distribution

A continuous random variable $X \in \mathbb{R}^+$ follows the log-logistic distribution, if and only if its logarithm $\ln X$ follows the logistic distribution [23], [21]. Note that, in our study, we deal exclusively with discrete random variables, as we discuss in section IV. The log-logistic Cumulative Distribution Function (CDF) is the sigmoid, as defined below:

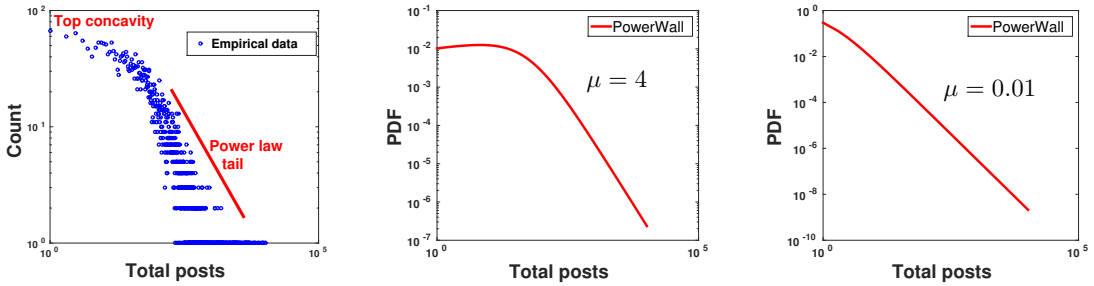
$$F_{LL}(x) \quad = \quad \frac{1}{1+e^{-w}} \qquad x \in \mathbb{R}^+ \tag{1}$$

where $w = \frac{\ln(x)-\ln(\mu)}{\sigma}$, $\sigma > 0$ is the shape parameter, $\mu > 0$ is the location parameter and $\rho = 1/\sigma$ is the slope of tail of the distribution.

A tell-tale property of the log-logistic is that its *Odds Ratio* plot is a power law. The odds-ratio function, $OR(t)$, is used in the survival analysis and is the ratio between the number of individuals that have not survived by time t versus the ones that survived. Formally:

Definition 1: The odds ratio function of a distribution is the ratio of CDF over the complementary CDF:

$$OR_{LL}(t) = \frac{F_{LL}(t)}{1 - F_{LL}(t)} \tag{2}$$



(a) PDF (b) *PowerWall* with $\mu = 4$ (c) *PowerWall* with $\mu = 0.01$

Fig. 2. **Suitability, and generality, of proposed *PowerWall***: (a) PDF of real data (blue dots - number of total posts on each user's wall for dataset D-12) - it exhibits "top concavity", *violating the power-law*. Our proposed *PowerWall* remedies that: being more general, it can exhibit top-concavity (see (b)), or not (see (c)), depending on the value of the location parameter μ.

More specifically, the Odds Ratio of a log-logistic distribution is linear on a double logarithmic plot with slope ρ and an intercept $-\rho \ln \mu$ (see, e.g., [10], or substitute Eq. 1 in the definition of odds ratio).

III. Dataset

In this section, we present in more detail the datasets and the feature set that we use to describe the behavior of Facebook users. The data was collected through a Facebook app that was voluntarily installed by the users. We opt to not refer to the app to provide an additional layer of protection for the users identity. We analyzed approximately 2.5 million posts of $7,310$ unique users spanning over four months in 2011, 2012 and 2013. Table II summarizes all of our datasets. There is an overlap of nearly 1000 users in the three datasets.

The dataset consists of activity of Facebook users who share information on their walls. The *wall* is an interactive feature in a user's Facebook profile where the user and her friends can interact through posts, likes or comments. There are three main types of posts: (i) plain text, (ii) media like photos and videos, or (iii) URL link. Each post belongs to one such category, and Facebook itself assigns a type, which we also use in our study. A *post* on a user's wall can be either a **self post**, provided by the user, or a **friend post** created by a friend of that user. A *comment* is a response to a post, and appears as such, and is categorized accordingly by Facebook as well. A *like* is essentially a non-verbal comment that shows support or appreciation for a post, and it is done through a distinct button that appears under every post.

TABLE II. Summary of Datasets

Dataset	Year	Period	# users	# posts
D-11	2011	Jul-Oct	5,170	1,159,647
D-12	2012	May-Aug	3,123	729,059
D-13	2013	Mar-Jun	2,779	537,881

For each post, we obtain a set of fields that describe the metadata of the post. Some of the fields include the original author of the post, time stamp of post creation, number of likes and comments, type of the post - status update, link, photo or other, application ID and name if the post was created via a Facebook application.

TABLE III. Features and Definitions

ID	Feature Name	Description
F_1	Total posts	# posts on users' wall
F_2	Self posts	# posts by user on her wall
F_3	Total likes	# likes on all posts
F_4	Self likes	# likes on self posts
F_5	Total comments	# comments on all posts
F_6	Self comments	# comments on self posts
F_7	Link posts	# self posts of type LINK
F_8	Photo posts	# self posts of type PHOTO
F_9	Status posts	# self posts of type STATUS

From the data described above, we choose a set of features that, we think, aptly describe the interaction between the user and her wall. We choose nine features of instructive per-user characteristics that measure the behavior of individual users in a wall-centric point of view. Table III list the features generated. Self-posts seem to dominate one's wall. We find that the posts from friends constitute anywhere between 11% and 17% to the total posts on a user's wall in the three datasets.

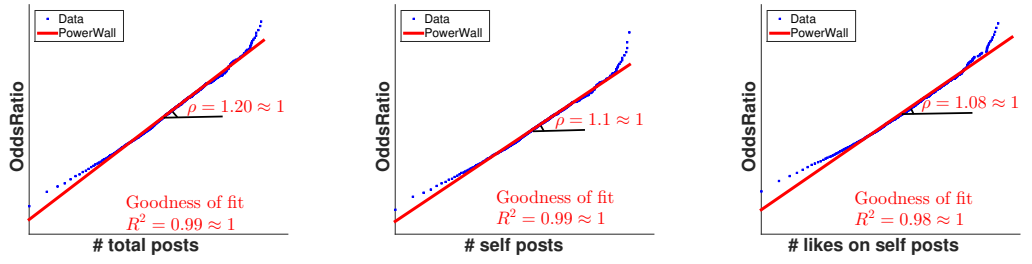
IV. Main Discovery: *PowerWall*

We show that *PowerWall*, which we define below, describes the empirical distributions for a select group of features that depict the behavior of Facebook users. In addition, we show that *PowerWall* includes power-laws as a special case, and observe that the slope parameter ρ is very close to 1, regardless of the time interval, or the population of users for our selected group of features.

A. What the data looks like

TABLE IV. Comparison of Distributions

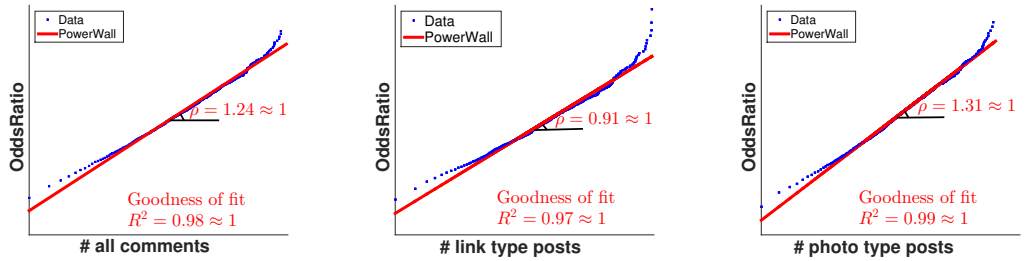
	Distribution name			
Property name	Pareto	Log-normal	Exponential	***PowerWall***
Skewed	✓	✓	✓	✓
Top-concavity		✓		✓
Power-law tail	✓			✓



(a) Odds Ratio of F_1 Count of total posts

(b) Odds Ratio of F_2 Count of self comments

(c) Odds Ratio of F_4 Count of likes on self posts



(e) Odds Ratio of F_5 Count of all comments

(f) Odds Ratio of F_7 Count of link-type self posts

(g) Odds Ratio of F_8 Count of photo-type self posts

Fig. 3. **Consistently good fitting of *PowerWall***: OddsRatio of the empirical distributions (in blue dots), vs *PowerWall*, in red line for features for the *D-12* dataset. All axis are logarithmic and The R^2 of the fit is over 0.95 and typically over 0.97 (with 1 being the ideal).

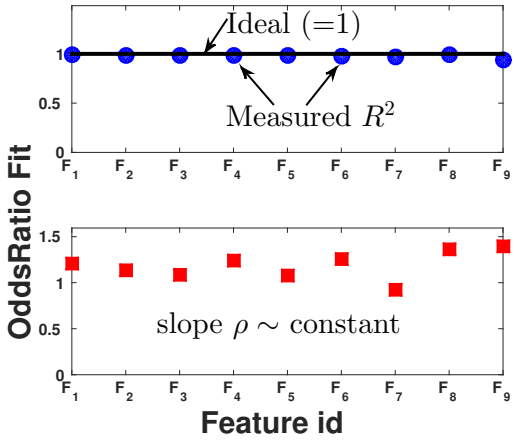


Fig. 4. **Persistence of *PowerWall* across features:** For all nine features for D-12: the approximation accuracy R^2 and the perfect fit represented by the line at $y = 1$, and the slope ρ which is close to 1.

We begin by presenting the challenge in modeling the observed user behaviors. Figure 2(a) shows the histogram of the total number of posts (F_1) on a user's wall for the *D-12* dataset. The distributions for all features were qualitatively similar across all three datasets. An indication of that is presented in Figure 3. An extensive table of all features, their distribution parameters, and their approximation is omitted due to space limitations.

We observe that the distribution has a heavy-tail and describing the distribution is not straightforward. It cannot be described by distributions with exponentially decaying tails. The distribution deviates significantly from a power law, which would have appeared as a straight line in this plot. At the same time, the Pareto distribution (a power-law probability distribution), which is also a heavy-tail distribution, lacks the flexibility to model a distribution that exhibits this flatness for low values (number of posts between 1 and 10 in our plot). Table IV compares distributions that have been widely used to model skewed distributions. The log-normal is a parabola on log-log scale, and if appropriately masked, may seem like a power-law. However, lognormal is more concentrated in the median and does not support as well the flatness for low values. This necessitates the use of a different model to accurately describe the observed distribution.

B. *The* PowerWall

We turn to the Odds Ratio function that we described earlier. In Figure 3, we plot the Odds Ratio for six of our metrics, which seems to be well approximated by a line in our logarithmic x and y axis.

For the Odds Ratio fitting, we perform a linear regression using Ordinary Least Squares (OLS) for all our empirical distributions. In addition, we also evaluate and verify the fitness using Maximum Likelihood Estimation (MLE). For a fitted polynomial of the form $a_0 \ln x + a_1$, the slope parameter is $\rho = a_0$ and $\mu = -\exp(a1/a_0)$. The Odds Ratio manages

to display the cumulative behavior in both the head and in the tail of the distribution at the same time.

We denote the above distribution of a discrete random variable to be *PowerWall*. Specifically, we define *PowerWall* as the distribution whose CDF is given below:

$$F_{PW}(x) = \frac{x^{\rho}}{x^{\rho} + \mu^{\rho}} \qquad x \in \mathbb{N}^{+} \tag{3}$$

where $\rho > 0$ is the slope parameter, $\mu > 0$ is the location parameter in following the notation used in the literature for consistency [23]. The distribution has an interesting property: its Odds Ratio $OR_{PW}(x)$ is linear on a double logarithmic scale with a slope ρ:

$$OR_{PW}(x) = \mu^{-\rho} x^{\rho} \tag{4}$$

$$\ln(OR_{PW}(x)) = \rho \ln x - \rho \ln \mu \tag{5}$$

Furthermore, most of the features we study have a linear approximation with a slope of roughly $\rho = 1$, as we see below. This similarity in slope makes *PowerWall* even more intriguing. As we saw in the previous section, earlier works have focused on variable random variables and they did not focus on a particular slope.

C. *Ubiquitousness:* PowerWall *describes many behaviors*

We find that all nine of our users behaviors captured by our features seem to described by a *PowerWall*. In other words, Odds Ratio for every feature in our set can be approximated by a line in double logarithmic plot. We use the *coefficient of determination*, also known as R^2, to evaluate the goodness of the fit. The closer R^2 is to 1, the better the model fits the data. Figure 3 displays the Odds Ratio distributions and their linear approximation for six features in the *D-12* dataset. There is a good fitness in all these plots with $R^2 > 0.97$.

In Figure 4, we plot the approximation accuracy R^2 and the slope ρ for all the nine features for year 2012, as captured in D-12. The approximation is quite strong and the slope is consistently close to 1. In fact, for all the nine features across the three years, the approximation holds with R^2 always above 0.95, except for the status self posts (F_9), which we discuss in section V. This pervasiveness suggests that *PowerWall* is a good model for the user behavior as seen through Facebook walls.

D. *Slope ρ: a surprising near-invariant*

The slope of the *PowerWall* Odds Ratio for the features in our dataset always lies in the range $0.9 - 1.2$. We have noticed that the ρ value for each feature remains fairly constant across all datasets, typically within 8% of its mean value. An indication of this stability is shown in Figure 5 for distribution of total posts (F_1). On the left plot, we see the stability of the ρ and R^2 values on the three different datasets across the three years. Intrigued by this, we wanted to further stress test this by looking at a weekly granularity. On the right plot, we separate dataset *D-12* in weeks, and plot R^2 and slope ρ for the feature total posts (F_1) for *D-12*.

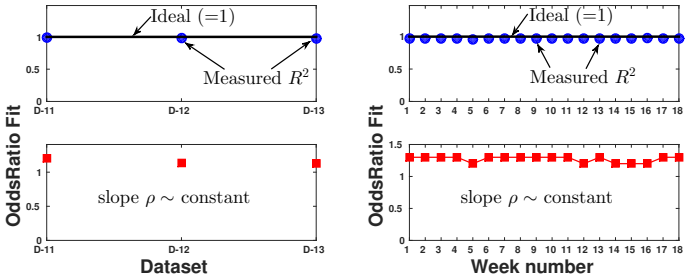


Fig. 5. **Persistence in time:** The invariance of the *PowerWall* for the feature total posts (F_1) reporting the approximation accuracy R^2 and the perfect fit represented by the line at $y = 1$, and the slope ρ which is close to 1 for: (a) across the three datasets, and (b) on a weekly basis in dataset *D-12*.

E. *Analysis - Pareto as a special case of* PowerWall

The *PowerWall* is heavy-tailed distribution that has Pareto tails on both ends. Pareto distribution is often invoked in connection with a long tail theory, but also with many other types of observable phenomena.

Lemma 1: A *PowerWall* distribution with parameters μ and ρ, includes Pareto as a special case, that is, asymptotically behaves like a a power law of slope $\alpha = 1 + \rho$.

Proof: Let X by a random variable in $(0, +\infty]$ that obeys the *PowerWall*, which means its CDF follows, Equation 3, with parameters μ and ρ. Differentiating the CDF, we obtain the PDF of the distribution below:

$$f_{PW}(x) = \frac{\rho(x/\mu)^{\rho}}{x(1 + (x/\mu)^{\rho})^2}$$

For $x \to \infty$, we have that $x \gg \mu$, $x/\mu \gg 1$ and

$$(1 + (x/\mu)^{\rho})^2 \simeq x^{2\rho}$$

and eventually:

$$f_{PW}(x) \propto x^{-(1+\rho)} \quad x \to \infty$$

which completes the proof. ■

The above lemma has the following implications that are captured in Figure 2: when the location parameter μ is small (say 0.001), *PowerWall* shows no top concavity, because $x \gg \mu$ for any positive integer value of x, and thus behaving like a Pareto with slope $\alpha = 1 + \rho$; when μ is large (say 4), then *PowerWall does* exhibit top-concavity, matching the real data.

V. Identifying Surprising Behaviors

As we saw, the observed *PowerWall* describes a large number of Facebook users behaviors, and the near-constant ρ is even more surprising. Can we put all these observations to practical use?

This is exactly what we attempt to do in this section. We provide initial evidence that, deviations from the *PowerWall* pattern are usually due to strange, and often robot-like behavior. Specifically, we study the top 5 outliers from status type posts (F_9) in *D-11* and *D-12*, which is one of

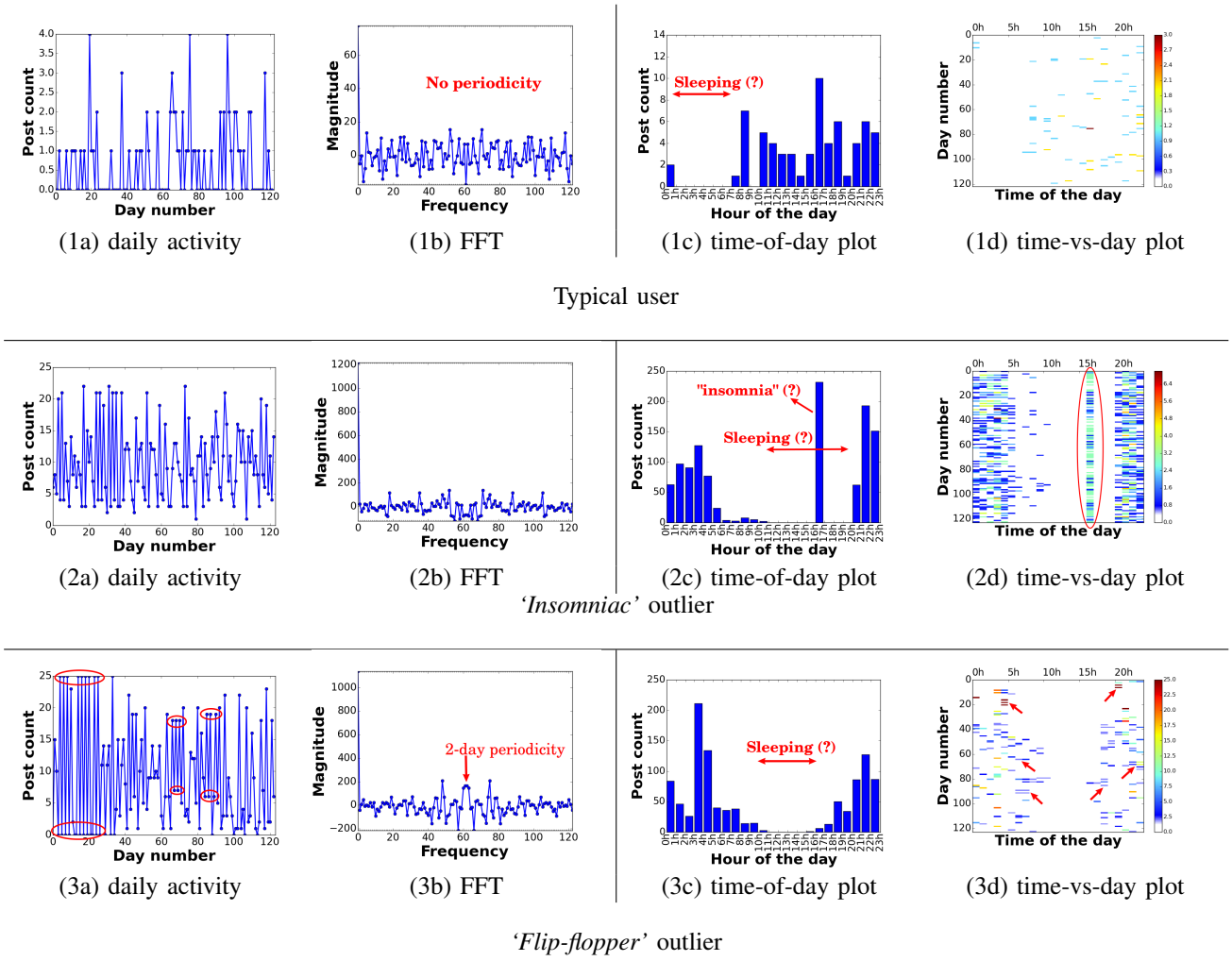


Fig. 6. ***PowerWall*** spots outliers: The top row represents the behavior of a typical user and the next two rows are for the *'Insomniac'* user and *'Flip-flopper'* user. *Middle row - the* 'Insomniac' *outlier*: consistent activity at 16:00 UTC, while apparently sleeping (no other activity before, or after, for several hours). Non-organic behavior, possibly due to a cron-job, or hijacked account. *Bottom row* - 'Flip-flopper' *outlier*: notice the wild 2-day oscillations (25 posts; then nothing; repeat).

the most deviating features, with low R^2 value. We present investigation results for two of the users that have more pronounced surprising behaviors.

Metrics for user-specific behavior investigation. We propose and generate a set of metrics and associated plots to study these outliers and discover some interesting and surprising behaviors. We showcase these plots in Figure 6:

1. Daily time series (1a-3a) presents the time series plot of the status posts made by the user per day.

2. Fast Fourier transform (FFT) of daily time series (1b-3b) to detect any periodicity.

3. Aggregated time plot (1c-3c) is the aggregate number of status posts made at a given hour during the period of observation. This shows the hours that the user is most active, and for example, help us infer her sleeping patterns. As we are not aware of the time zone of of the user, we use the given Greenwich time, which most likely does not coincide with the local time of the user.

4. Time-day plot (1d-3d) provides a heat-map of the number of user posts behavior per hour (x-axis) and per day (y-axis).

Here, we focus on the status type of posts but the same plots could be generated for our other features to cast a wider net for capturing surprising behaviors.

A. *Typical user*

To highlight the weirdness of the behavior of the users we spotted, we also show the plots for a 'typical' user - which we selected among the users that do obey the *PowerWall*; this user has 60 status posts, and comes from the *D-13* dataset. Figure 6 (1a) to (1d) shows the behavior of that typical user Figure 6 (1a) and (1b) show the daily activity and its Fourier Transform: notice the lack of periodicity - not even a weekly

one. This is probably due to smartphones: people don't need to be in front of a desktop, to post on facebook. Figure 6 (1c) is UTC time-of-day plot - notice that there is a gap of 7-9 hours of inactivity (red double-arrow), obviously the sleeping time for that user in his/her timezone. This is corroborated by the 'time-vs-day' plot in Figure 6 (1d): there are no posts between 01h and 07h (UTC), throughout the full four month duration of the dataset. This means that our user maintained his/her sleeping pattern, and probably he/she did not travel outside the original timezone.

In short, we would expect 'typical' users to have (a) some fairly consistent sleeping interval and (b) no periodicities. The two outliers that *PowerWall* spotted, each violate these expectations, as we discuss next.

B. 'Insomniac' *outlier*

We analyze the behavior of the most active user in terms of status posts, who deviates from the *PowerWall*. Figure 6 (2a)-(2d) shows the status posting behavior of that user. The most interesting plot is Figure 6 (2c) and (2d), which show that there is a single hour with high posting activity within a quiet period with zero posts. If we conjecture that the quiet time is the night time, the user wakes up at exactly the same time almost every night and creates status posts. Upon investigation, we find that all status posts, including the ones in the "isolated" hour, are generated through Twitter's Facebook application.

C. 'Flip-flopper' *outlier*

We also find that the third most active user in terms of status posts exhibits surprising behavior, which we study in Figure 6 (3a)-(3d). In this case, the surprising behavior is a very precise oscillating posting behavior, which is best captured in Figure 6 (3a). The user alternates between zero and 25 posts per day as highlighted by the red circles. Furthermore, this kind of oscillations appear in subsequent periods, such as between day 60 and 80 with daily status posts that go from seven to 19. This is the reason that the FFT plot has a spike at frequency 60 and 61, which correspond to period T=2 days (see red arrow in Figure 6(3b)). The spike at 47 probably corresponds to period 3, when the user skips-a-beat. We accentuate these oscillations in Figure 6(3d) as well. Upon investigation, the oscillating behavior is due to *dlvr.it*, which is a Facebook app that enables and automates sharing of news and information. In fact, we find that 78% of all the status posts were done via this app!

VI. Related work

There is limited work on modeling the posting behavior of Facebook users, which is our focus here. We briefly review related work, which falls into the following research directions.

Facebook user activity profiling. A lot of work has been done in identifying user preferences and traits by profiling the user's online social activity. Kosinksi et al. [24] used Facebook 'Likes' to predict user traits like age, gender, ethnicity, political and religious affiliations etc. [5] proved the existence of Dunbar's number in Facebook ego-networks. Dunbar's number, 150, defines the number of individuals with whom it is possible to maintain stable interpersonal relationships [25].

Extracting information from user activity. Wang et al.[14] studied how social and smartphone activity of students correlates to changes in their life throughout the semester. Similarly, De Choudhury et al. tried to infer pathological conditions such as postpartum depression [13].

Psychological and social studies. Acquisti and Gross [26] studied user attitudes towards privacy in Facebook. Gosling et al. made qualitative observations on human trends and behaviors using Facebook data [27].

Modeling Facebook friendship network. Several studies attempt in modeling the macroscopic properties of the Facebook friendship network [1], [28]. These studies found that, unlike other topologies of complex systems, the Facebook friendship graph cannot be accurately modeled by a power law. Other efforts focus on communities, ego-nets and their evolution in Facebook [29], and popular blog networks [4]. Several works have studied the tie strength between Facebook users based on their interactions between users and their friends [30]. Note that none of these works focus on modeling the posting behavior of Facebook users.

Other social networks and communication platforms. Finally, there is a plethora of studies on other social networks, most notably Twitter, and various communications mediums, such as the telephone call graphs and calling patterns for both wired and cellular networks, text messages etc [11], [31], [32]. In addition, Twitter has received significant more attention due to its data accessibility. Note however, that Twitter is predominantly a broadcast mechanism, and thus a significantly different medium.

VII. Conclusions

We studied the behavior of real users, posting on Facebook walls. Our key contributions are:

1) **Distribution - *PowerWall***: Many behaviors, such as the number of posts and comments, obey our *PowerWall* with consistently close fit (R^2 over 0.95).
2) **Pervasiveness of *PowerWall***: The slope of the Odds Ratio plot is near-invariant, with value $\sim$1 (from 0.9 to 1.2). Moreover, these properties hold not only for many different features, but also for: (a) a span of three years, (b) three largely different user groups, and (c) for two different time granularities, a four-month and weekly intervals.
3) **Usefulness**: Deviations from *PowerWall* point to strange users, who, upon investigation, exhibit non-organic/script-like behavior. Thus, our model can be used for outlier detection.

Reproducibility: We have open-sourced our code, at *https://github.com/pdevineni/powerwall/*, and we will share as much of the data as we can (subject to user-privacy considerations).

Acknowledgements

This material is based upon work supported by the National Science Foundation under Grants No. SaTC 1314935, IIS-1217559, CNS-1314632, IIS-1408924, by DARPA grant Social

Media in Strategic Communication (SMISC) program under Agreement Number W911NF-12-C-0028 and by the Army Research Laboratory under Cooperative Agreement Number W911NF-09-2-0053.

Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation, or other funding parties. The U.S. Government is authorized to reproduce and distribute reprints for Government purposes notwithstanding any copyright notation here on.

REFERENCES

[1] M. Gjoka, M. Kurant, C. T. Butts, and A. Markopoulou, "Walking in facebook: A case study of unbiased sampling of osns," in *Proceedings of the 29th Conference on Information Communications*, ser. INFOCOM'10. Piscataway, NJ, USA: IEEE Press, 2010, pp. 2498–2506.

[2] B. Viswanath, A. Mislove, M. Cha, and K. P. Gummadi, "On the evolution of user interaction in facebook," in *Proceedings of the 2Nd ACM Workshop on Online Social Networks*, ser. WOSN '09. New York, NY, USA: ACM, 2009, pp. 37–42.

[3] P. Boldi and S. Vigna, "Four degrees of separation, really," in *Proceedings of the 2012 International Conference on Advances in Social Networks Analysis and Mining (ASONAM 2012)*, ser. ASONAM '12. Washington, DC, USA: IEEE Computer Society, 2012, pp. 1222–1227.

[4] S. Gupta, X. Yan, and K. Lerman, "Structural properties of ego networks," in *International Conference on Social Computing, Behavioral Modeling and Prediction*, April 2015.

[5] V. Arnaboldi, A. Passarella, M. Tesconi, and D. Gazzè, "Towards a characterization of egocentric networks in online social networks," in *Proceedings of the 2011th Confederated International Conference on On the Move to Meaningful Internet Systems*, ser. OTM'11. Berlin, Heidelberg: Springer-Verlag, 2011, pp. 524–533.

[6] T.-K. Huang, M. S. Rahman, H. V. Madhyastha, M. Faloutsos, and B. Ribeiro, "An analysis of socware cascades in online social networks," in *Proceedings of the 22Nd International Conference on World Wide Web*, ser. WWW '13. Republic and Canton of Geneva, Switzerland: International World Wide Web Conferences Steering Committee, 2013, pp. 619–630.

[7] J. Cheng, L. Adamic, P. A. Dow, J. M. Kleinberg, and J. Leskovec, "Can cascades be predicted?" in *Proceedings of the 23rd International Conference on World Wide Web*, ser. WWW '14. New York, NY, USA: ACM, 2014, pp. 925–936.

[8] B. Gonccalves, N. Perra, and A. Vespignani, "Modeling users' activity on twitter networks: Validation of dunbar's number," *PloS one*, vol. 6, no. 8, p. e22656, 2011.

[9] A. A. Nanavati, S. Gurumurthy, G. Das, D. Chakraborty, K. Dasgupta, S. Mukherjea, and A. Joshi, "On the structural properties of massive telecom call graphs: Findings and implications," in *CIKM'06*, 2006, pp. 435–444.

[10] P. O. S. V. De Melo, L. Akoglu, C. Faloutsos, and A. A. F. Loureiro, "Surprising patterns for the call duration distribution of mobile phone users," in *ECML PKDD'10*, 2010, pp. 354–369.

[11] D. Koutra, V. Koutras, B. A. Prakash, and C. Faloutsos, "Patterns amongst Competing Task Frequencies: Super-Linearities, and the Almond-DG Model," in *PAKDD*, 2013, pp. 201–212.

[12] L. Guo, E. Tan, S. Chen, X. Zhang, and Y. E. Zhao, "Analyzing patterns of user content generation in online social networks," in *Proceedings of the 15th ACM SIGKDD international conference on Knowledge discovery and data mining*. ACM, 2009, pp. 369–378.

[13] M. De Choudhury, S. Counts, E. J. Horvitz, and A. Hoff, "Characterizing and predicting postpartum depression from shared facebook data," in *Proceedings of the 17th ACM Conference on Computer Supported Cooperative Work & Social Computing*, ser. CSCW '14. New York, NY, USA: ACM, 2014, pp. 626–638.

[14] R. Wang, F. Chen, Z. Chen, T. Li, G. Harari, S. Tignor, X. Zhou, D. Ben-Zeev, and A. T. Campbell, "Studentlife: assessing mental health, academic performance and behavioral trends of college students using smartphones," in *Proceedings of the 2014 ACM International Joint Conference on Pervasive and Ubiquitous Computing*. ACM, 2014, pp. 3–14.

[15] M. Faloutsos, P. Faloutsos, and C. Faloutsos, "On power-law relationships of the internet topology," *SIGCOMM*, pp. 251–262, Aug-Sept. 1999.

[16] A. L. Barabási and R. Albert, "Emergence of scaling in random networks," *Science*, vol. 286, pp. 509–512, 1999.

[17] S. Redner, "How popular is your paper? an empirical study of the citation distribution," *European Physical Journal B*, vol. 4, no. 2, pp. 131–134, Aug. 1998.

[18] A. Clauset, C. R. Shalizi, and M. E. J. Newman, "Power-law distributions in empirical data," *SIAM Rev.*, vol. 51, no. 4, pp. 661–703, Nov. 2009.

[19] M. Mitzenmacher, "A brief history of generative models for power law and lognormal distributions," *Internet Mathematics*, vol. 1, no. 2, pp. 226–251, 2004.

[20] M. Seshadri, S. Machiraju, A. Sridharan, J. Bolot, C. Faloutsos, and J. Leskovec, "Mobile call graphs: beyond power-law and lognormal distributions," in *Proceeding of the 14th ACM SIGKDD international conference on Knowledge discovery and data mining*. Las Vegas, Nevada, USA: ACM, 2008, pp. 596–604.

[21] P. O. S. Vaz de Melo, C. Faloutsos, R. Assunção, and A. Loureiro, "The self-feeding process: A unifying model for communication dynamics in the web," in *Proceedings of the 22Nd International Conference on World Wide Web*, ser. WWW '13. Republic and Canton of Geneva, Switzerland: International World Wide Web Conferences Steering Committee, 2013, pp. 1319–1330.

[22] J. I. McCool, *Using the Weibull distribution: reliability, modeling and inference*, ser. Wiley Series in Probability and Statistics. Hoboken, NJ: Wiley, 2012.

[23] D. G. Champernowne, "The graduation of income distributions," *Econometrica: Journal of the Econometric Society*, pp. 591–615, 1952.

[24] M. Kosinski, D. Stillwell, and T. Graepel, "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, 2013.

[25] R. I. Dunbar, "Neocortex size as a constraint on group size in primates," *Journal of Human Evolution*, vol. 22, no. 6, pp. 469–493, 1992.

[26] A. Acquisti and R. Gross, "Imagined communities: Awareness, information sharing, and privacy on the facebook," in *Proceedings of the 6th International Conference on Privacy Enhancing Technologies*, ser. PET'06. Berlin, Heidelberg: Springer-Verlag, 2006, pp. 36–58.

[27] S. D. Gosling, S. Gaddis, S. Vazire *et al.*, "Personality impressions based on facebook profiles." *ICWSM*, vol. 7, pp. 1–4, 2007.

[28] J. Ugander, B. Karrer, L. Backstrom, and C. Marlow, "The anatomy of the facebook social graph," *CoRR*, vol. abs/1111.4503, 2011.

[29] L. Backstrom, D. Huttenlocher, J. Kleinberg, and X. Lan, "Group formation in large social networks: membership, growth, and evolution," in *Proceedings of the 12th ACM SIGKDD international conference on Knowledge discovery and data mining*. ACM, 2006, pp. 44–54.

[30] E. Gilbert and K. Karahalios, "Predicting tie strength with social media," in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, ser. CHI '09. New York, NY, USA: ACM, 2009, pp. 211–220.

[31] T. Karagiannis, J.-Y. Le Boudec, and M. Vojnovic, "Power law and exponential decay of intercontact times between mobile devices," *Mobile Computing, IEEE Transactions on*, vol. 9, no. 10, pp. 1377–1390, 2010.

[32] A. Mislove, M. Marcon, K. P. Gummadi, P. Druschel, and B. Bhattacharjee, "Measurement and analysis of online social networks," in *Proceedings of the 7th ACM SIGCOMM Conference on Internet Measurement*, ser. IMC '07. New York, NY, USA: ACM, 2007, pp. 29–42.