Predicting Fitness Behavior Based On Online Social Interactions

Alice Yu, Mihail Eric, Geet Sethi

In this work we explore the effect of a user's online social interactions on their offline fitness behavior. Using data from the fitness-focused social media app Fitocracy, we extract various features ranging from social network structure, sentiment analysis of user interactions, and the evolution of workout data to build a model which can predict, based on the aforementioned features, whether joining a group can positively or negatively affect a user's offline fitness activity. To our best knowledge, this is first model of its kind using data from the Fitocracy site, and it is able to predict user workout changes in fitness communities with a 70% f_1.

1 Introduction

Exercise is becoming increasingly important in today's society. The increase in societal awareness of health risks combined with the growth of social media have led to the production of many mobile fitness apps. Users of these apps can track their exercises, receive feedback, and record their daily fitness activity. However, it is difficult to sustain user engagement by simply asking users to track their progress without any extrinsic rewards or social structures in place. To address this, creators of the Fitocracy fitness app decided to gamify fitness. Fitocracy channels people's competitive nature by allowing users to complete challenges and compete against others when they complete more workouts. By recording user workouts, Fitocracy can track offline social activities. It also provides a social platform for people to communicate about fitness and make online social fitness connections. This plethora of recorded workout information and user-user interactions forms an interesting testing bed for understanding the connection between user's online behaviors and their offline activities.

1.1 Problem Statement

The question we wished to answer in our project is whether or not it is possible to predict a user's offline behavior, in this case their fitness activity, based on their activity in the Fitocracy social network. To do so, we investigated three major areas of the social network and dataset: 1) the network dynamics and structure between users and groups; 2) sentiment analysis of the textual content that users post amongst themselves and to groups; 3) the structure and evolution of the workouts completed by individual users and across groups.

2 Prior Work

Previous research has touched upon understanding the online social dynamics in the context of fitness and incorporating it with exercise measurements. However, this information is gained independently, and it is hard to deduce the immediate effects of the social interaction with a user's workout behavior. Cunha et al., proposed a method to measure whether positive reinforcement increases a user's return to one of Reddit's subreddit loseit in the future [6]. They created a topical representation of the first post's content extracted using Latent Dirichlet Allocation (LDA). Although the study analyzed the communication dynamic within fitness posts, they could only measure a user's fitness habits due to comments based on context clues within user comments. Adar et al. also performed large scale studies attempting to formalize the correlation of users in time-varying datasets and their behavioral properties in order to predict patterns [13]. With the Fitocracy dataset, we can address the issue present in Cunha et al. because we have user workout information; and we explore the ideas presented by Adar et al. in a fitness setting due to the prevalence of timestamped information in the Fitocracy dataset.

Althoff et al. studied the effects of social networking features in a fitness tracking application to try and understand how user behavior, both online and offline, is impacted [7]. However, in their analysis of the dataset and quantification of the impact of social influence on the social networks users, the authors only focused on the relationship between users, that is, the friendship graph between users. The authors missed out on exploring the depth behind, and influence of, the interactions among the users of the social network as relationships were reduced to simply an edge. These relationships, which can take the form of users posting comments of encouragement to other users, can drastically impact one's experience and affect both their online and offline behavior. The lack of this kind of analysis is one of the driving forces behind our choice to study the community dynamics in the Fitocracy fitness network. We analyzed the Fitocracy social network structure by incorporating friendship clusters, network sizes, textual analysis, and offline activities information.

3 Background

3.1 Data

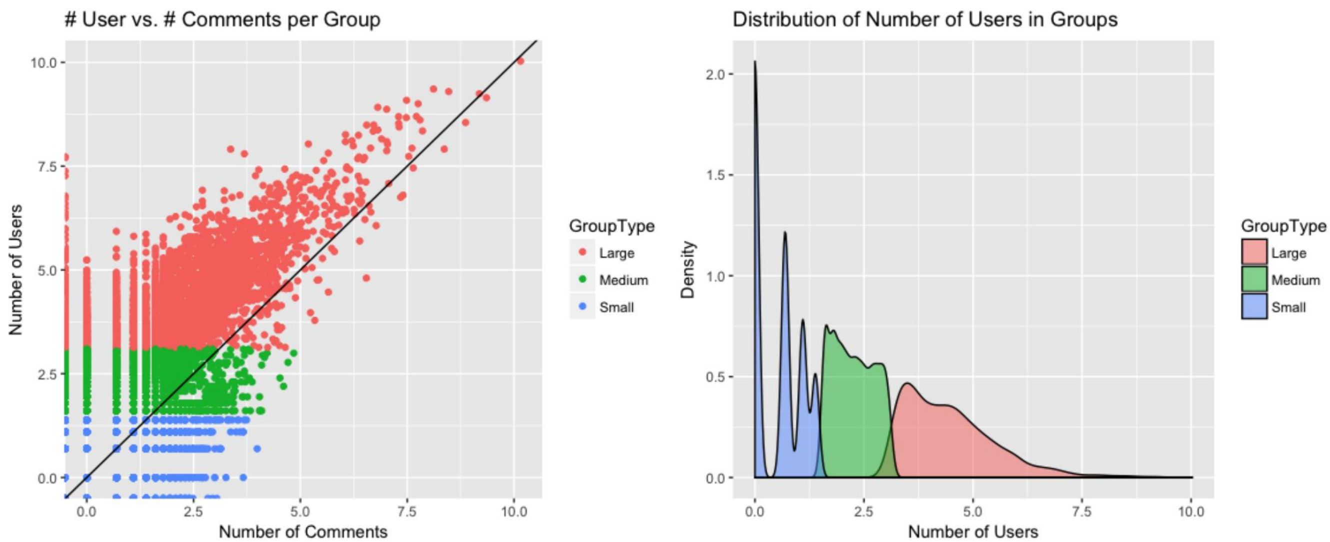
The data we are proposing to investigate was collected by Jurgens et. al. [5] and consists of the complete workouts and profiles of 441,034 users, recording historical activity over the period February 2011 to January 2015. The data includes a number of group statistics such as number of members and average fitness level (as designated by the Fitocracy site), the full social network of users, which is directed like that of Twitter, the group join dates of each user, user metadata such as age, gender, level, and all the question/answer/comment threads posted by a user to their own wall and to the groups they have joined.

The gamification of user fitness goals and benchmarks done by Fitocracy through its point/level system presents an interesting element underlying the social and community dynamics of the platform. We are fundamentally interested in seeing how this level-based system plays out in the domain of fitness and how we can understand the interplay between users'fitness behaviors and their interactions with other users, including through text-based interactions. We are performing the first large-scale investigation of the Fitocracy dataset at the level of user-user interactions.

3.2 Network Representation

Our data was represented in several graphs. The first graph was a directed graph where each node represents a unique user or group, and edges represents the relationship between two nodes. For example, Fitocracy's social network acts similar to that of Twitter, where users can follow one another (without the requirement of the followed reciprocating), hence why the graph was directed. Also, when joining a group, a bidirectional edge was drawn between the user and group as information flows from both the user to the group and from the group to each user. The second graph was a weighted multi-graph to model workouts, where a node represents either a unique user or exercise, and edges are drawn from a user to each exercise they complete, with the edge weight being the number of times the user did that particular exercise. The third graph was a directed multi-graph of social interactions once again, but in this graph nodes represent users and questions, where each question has a unique node, and asking/answering the question are the edges.

3.3 Community Definition



(a) User vs. Comments Per Group

(b) User Distribution

Figure 1: The plots above show various group statistics on a log-log scale

The Fitocracy website gives user the ability to create their own groups, intended for users with similar fitness interests to gather and discuss their workouts and goals. For our study, we augmented this Fitocracy-defined notion of communities to partition all such fitness groups into separate categories based on percentiles of number of users in these groups. This gave us three clusters of groups, which we are defining as *small*, *medium*, and *large*. The *small* cluster of groups, consists of all groups with 22 or fewer users, representing the 50th percentile and below of group sizes. The *medium* cluster of groups consists of all groups with between 4 and 22 users, representing the 50th to 75th percentile. The *large* cluster of groups consists of all groups with more than 22 users, representing all group sizes above the 75th percentile. Figure 1 shows a visual representation of how the group categories were defined. All of our analyses are done with respect to this number-of-users definition of communities. As seen in the left plot of Figure 1, the group types are clustered even when we are looking at different features of the network components, indicative that these clusters can be used for further analysis.

3.4 Preliminary Analysis

Below we provide general summary statistics of the dataset:

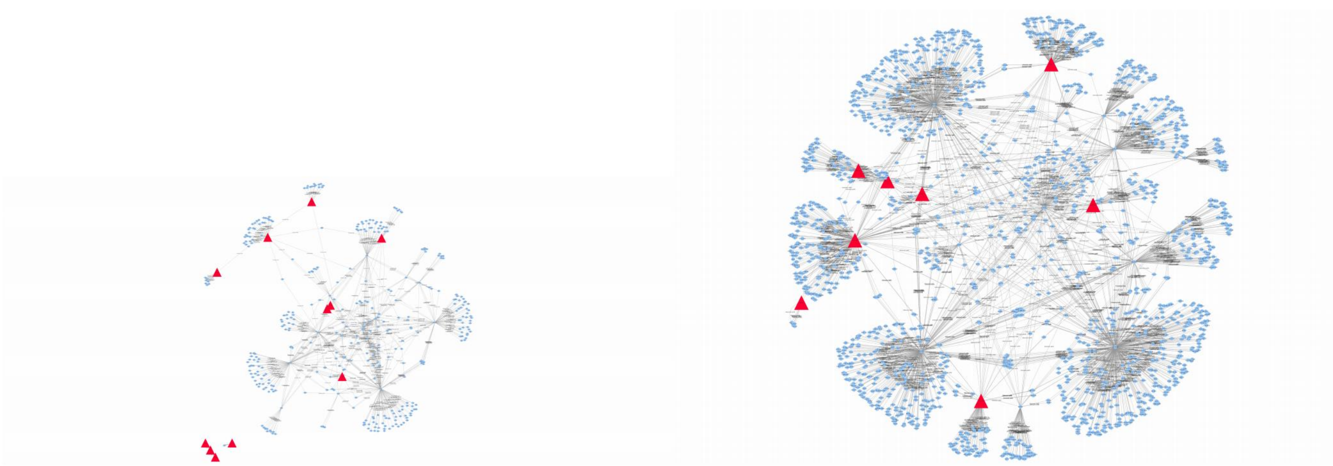
Number Users	53,508
Number Workouts	54,938,207
Number Edges Social Network Graph	3,311,204
Number Questions	193,717
Number User Wall Questions	131,748
Number Group Wall Questions	61,969
Number Groups	12,410

The size of each network component displayed in the table above proves that we have a large enough dataset to perform our following analyses.

As well as per user and group statistics:

Average Number of Edges per group	361.6
Average Number of Users per group	55.1
Average Number of Comments per group	16.8
Average Number of Workouts per user[1]	278

The second table outlines the properties of the groups we are analyzing. As seen in the average number of edges and users, we have relatively large groups, considering how many of these groups overlap in nodes and edges. Similarly, the average number of comments and workouts provided enough content for us to perform our semantic and user workout analyses.



(a) Group with low average degree centrality

(b) Group with high average degree centrality

Figure 2: Networks with varying average degree centralities

To find properties of the groups, we looked at degree centrality, betweenness centrality, and closeness centrality. We opted for degree centrality because we only wanted to measure how a simple online connection can relate to workout behavior. Figure 2 represents groups with the same number of people, but showing how connected to other members in the network each group member is. In the network representation of the group with low average degree centrality, we see how group members are either isolated from other Fitocracy members or connected with very different clusters of outside members. On the other hand, we see that in the network representation of a high average

[1] Over non-disqualified users, see §4.4

degree centrality group, the users are connected to a large number of people in other groups and they share similar outside connections. This suggests that the members of that community must actively seek online connections to motivate their offline behavior.

4 Methodology

Given the preliminary analyses above, we build off the hypothesis that the group dynamics that result from user-initiated, common-interest-driven communities can lead to an understanding of user offline activity. In particular, we can analyze characteristics of these communities in the following ways: 1) We can extract mathematical network properties, treating the communities as induced subgraphs of the larger Fitocracy network. 2) The communities form cohorts of users with online interactions represented by their textual interactions, which lend these communities to *softer* natural language analyses. We would like to combine these two approaches to explain offline activity represented through their workout behaviors. We do this by extracting features from these communities and building a predictive model of the tendency of a communities' users to see changes in their numbers of completed workouts over the lifetime of the users' activity on Fitocracy.

4.1 Model Description

We framed our problem as a binary classification where each data point was a single group on Fitocracy from which we extracted relevant features, and we predicted the aggregated delta of the number of workouts completed by users in a group, where a +1 indicated that the total number of workouts completed by users saw a positive delta over the lifetime of users on Fitocracy and a -1 indicated a negative or zero delta. We used a max-margin algorithm in the form of a support vector machine (SVM) [11]. The SVM was further augmented with a Gaussian kernel function that allows the model to learn nonlinear decision boundaries of the data. When selecting the modeling algorithm, we experimented with using a linear kernel function for the SVM as well as using a logistic regression classifier, though performance was best for the SVM with a Gaussian kernel.

4.2 Network Features

We computed the following network features for each induced sub-graph representing a group:

- **Number of Nodes**: Each node represents a user, and hence this feature gives some sense for the overall size of the group.
- **Number of Edges**: This feature counts the number of edges among users in the group and hence gives a coarse-grained measure of the number of connections between users.
- **Average Degree Centrality**: The average degree centrality measures the relative importance of each user in the group in the overall Fitocracy network. Examples of how this metric affects group clustering and network influence of users in the group are described above and shown in Figure 2.
- **Modularity**: The modularity is a measure of the presence of discrete communities within a group, calculated using the Clauset-Newman-Moore community detection algorithm. We opted for Clauset-Newman-Moore instead of the Girvan-Newman algorithm because of our large network size. This score acts as our measurement of friendship levels within a group compared to friendship levels with users outside of a group. If modularity is high within the

group, that means users are closer to members of the group. If users are not connected to members in the group, but with other Fitocracy members, then this may indicate the user is not in the group to make friends and receive fitness support.

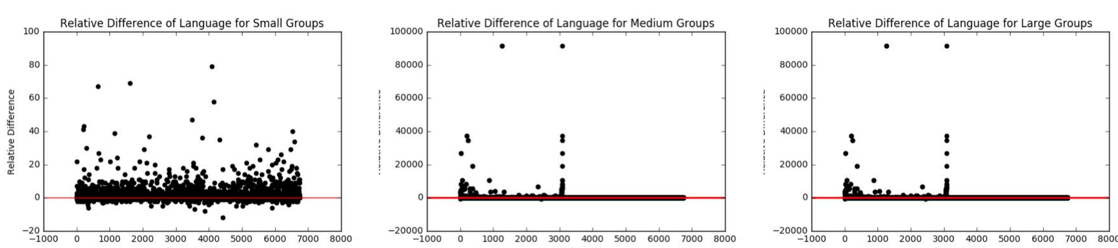
4.3 Natural Language Features

In order to characterize the nature of text-based interactions between users from different communities within the network, we performed sentiment analysis of all the comments by users in a group. To gauge sentiment in text, we used curated lexicons of tone-specific terms. We used two separate lexicons: one for positive words and one for negative words, both of which were developed by Hu et. al. Given that the Fitocracy textual exchanges are done in an online setting where vernacular for describing certain sentiments tends to be different, we augmented the positive/negative lexicons from Hu et. al with those of Wang et. al., which were scraped from Urban dictionary[9].

First, we attempted to form a low-dimensional lexical representation of the textual content of user posts, using Latent Dirichlet Allocation (LDA)[10]. This unsupervised algorithm seeks to determine the implicit topics of a collection of documents using a generative statistical model based on Dirichlet sampling. Here the documents we fed into the model were all the posts and comments made by users of particular group, and hence we were aiming to develop group-specific topics that would hold some lexical signal of a particularly positive or negative tone.

The words per topics were often not particularly meaningful, even when we tuned the number of topics extracted. Hence, running LDA often did not give an easily discernable signal for tone-specific expressions.

Since these low-dimensional representations of the text were not successful, we opted instead for a bag-of-words representation for each document, in order to classify their positive or negative sentiment. First, we processed the questions by removing punctuation and removing questions and comments below a certain length threshold because it is more difficult to characterize the emotion of very short utterances. Next, we used the lexicon lists from above and calculated the number of positive and negative words that were used in each question. These positive and negative word counts formed two features that were considered for the model. In addition, we calculated a relative tone score as a difference of the positive word count and the negative word count. Relative tone scores per comment of the different group sizes are shown in Figure 3.



(a) Small Sized Groups (b) Medium Sized Groups (c) Large Sized Groups

Figure 3: Each graph displays the question score calculated per user question in each network cohort.

4.4 Workout

The main feature we wished to extract from the workout network was the change in the number of exercises completed by the users in a particular group over their time on Fitocracy. To do so, we first filtered out users who would not be statistically significant based on their workout data. Features we filtered by were the number of days active on Fitocracy (i.e. users who were only active one day cannot be used to compute a delta) as well as the total length of time a user was active on Fitocracy (we required users to be active for at least one month to be included in our study).

$$\text{Consider group G} = \{u_1, u_2, ..., u_k\}$$
$$\text{score(G)} = \sum_{i \in G} \sum_{t=2}^{d_i} \delta_{t,t-1}^{i}$$
where $\delta_{t,t-1}^{i}$ denotes the workout delta of user i from day t-1 to day t

Equation 1: Algorithm for creating workout scores for each group

Once we filtered out the "disqualified" users, we then computed the absolute delta for each user, by tracking their daily change in workout activity over the course of their time on Fitocracy. This results in each user being assigned a positive or negative score representing the net change in their workout behavior while active on the social network. Once we have computed the scores for each user, we then compute a per group score by summing up all of the scores of the users within that group as seen in Equation 1. This represents the net change in workouts completed amongst all users in the group over their time on Fitocracy.

5 Results

In total, we generated features and labels for 12,552 distinct groups and used a 0.8/0.1/0.1 train-validation-test data split. The SVM was trained on the training data split, hyperparameters were tuned and features selected using the validation data, and then the best performing model and feature set were evaluated on the test set. The SVM used an l_1 error penalty coefficient of 0.01 to regularize the model and prevent overfitting. Evaluation was reported on an f_1 metric. The best performing model used a feature set consisting of edge count, node count, and average degree centrality. It produced the following scores for f_1, precision, and recall on the test set: **(0.70, 0.61, 0.89)**.

5.1 Discussion

To assess the expressiveness of the various features we extracted, below we include results from an ablation study done on subsets of the features. In particular, we ran an ablation by evaluating a model on the validation split with only a single feature from our feature set described in Section 4. The results on (f_1, precision, recall) are shown below:

Feature	Performance
Number of Nodes	(0.66, 0.65, 0.65)
Number of Edges	(0.66, 0.65, 0.66)
Modularity	(0.5, 0.61, 0.38)
Average Degree Centrality	(0.7, 0.6, 0.9)
Count of Positive Words	(0.61, 0.62, 0.59)
Count of Negative Words	(0.57, 0.62, 0.49)

In general, our results indicate that certain aspects of the induced mathematical network representing a group can be quite powerful in predicting the tendency of a group's users to increase or monotonically decrease their number of workouts over time. Though our labels for a datapoint use a fairly coarse-grained notion of changes in workout behavior, we believe they still can give a general sense for the workout behavior of a collection of individuals.

Our ablation study on the features produced some very interesting findings. For example, to our surprise, using softer natural language features derived from users' interactions with each other on Fitocracy do not have as much of an influence on user activity as we had expected. Using the sentiment analysis features from our curated lexicons do not perform particularly well on our predictive task. This could be because the lexicons alone do not capture enough of the nature of textual interactions on Fitocracy, and the signal from these features may not differentiate enough among group behaviors.

The number of nodes and number of edges in these induced subgraphs, on the other hand, were quite powerful indicators for a group's behavior. In general, these features capture the intuition that being in group's of different sizes may have an effect on a user's desire to change their workout behavior over the span of their activity on Fitocracy. The number of edges in a group is a coarse notion of connectivity among users which can be seen as a measure of the level of interaction among users within a group. Positive encouragement from a user with especially impressive fitness credentials can certainly influence a new user's desire to workout, and vice versa for negative interactions.

The most expressive feature from the study was the average degree centrality of the users in a group. This intuitively makes sense because in network terminology, the degree centrality of a node is a measure of that node's importance in a network. In sites with a social network component such as Fitocracy, this can directly translate to a node's ability to influence its surrounding nodes. Hence a group with a higher percentage of influential users can certainly impact the behavior of other users in that group. This is reflected by the expressive capacity of this particular subgraph feature.

6 Conclusion and Future Work

In this project, we have performed the first large-scale study of the Fitocracy dataset attempting to correlate the social interactions and community dynamics of online user activity with offline behavior. We did this by running a variety of analyses including natural language processing on user posts and mathematical analyses of network structure. This allowed us to build a statistical model using features of fitness groups on the site to predict the tendency of a group's users to increase or decrease their number of workouts over time, with a best-performing model f_1 of 70%.

There are a number of future directions we can pursue to improve the performance of our model. First off, we can use a more fine-grained notion of workout deltas over time as the labels our model is trying to predict. This may involve a more specific temporal framing of a user's activity and is perhaps better represented as a regression rather than classification problem. Secondly, we can continue to investigate different types of features. While our analysis did not suggest that sentiment analysis of comments was very expressive, it may be the case that we simply did not have enough lexical features to form a strong enough signal of user behavior. We could augment natural language features with unigram or bigram word counts of tokens in group posts or do more sophisticated semantic analysis through frame-based parsing.

A model such as ours could be integrated into recommendation systems for suggesting fitness groups new users should join. The fact that we were able to create a substantially predictive model of user workout behavior through network dynamics and community properties suggests that this flavor of analysis can be fruitful for websites such as Fitocracy that seek to positively impact individuals' health choices and overall well-being.

7 Acknowledgements

We are grateful to David Jurgens for his guidance in the development of our experimental design and for allowing us access to the Fitocracy data, and to Professor Leskovec and the whole CS 224W teaching team for their outstanding instruction and support throughout the quarter.

Author Contributions

Alice: project formulation and design, dataset and network analysis, network feature extraction and analysis, plotting of statistics and graphs, poster/paper writeup

Mihail: project formulation and design, utilities/readers for handling data, natural language analysis, model building/experiments/studies, poster/paper writeup

Geet: group/user workout studies/features, edits on paper

References

[1] "Danescu-Niculescu-Mizil et al., How Opinions are Received by Online Communities: A Case Study on Amazon.com Helpfulness Votes, WWW 2009 <http://snap.stanford.edu/class/cs224w-readings/Danescu09OpinionHelpfulness.pdf >.

[2] "Anderson et al., Discovering Value from Community Activity on Focused Question Answering Sites: A Case Study of Stack Overflow, ACM SIGKDD international conference on Knowledge discovery and data mining, 2012 <http://snap.stanford.edu/class/cs224w-readings/sof-kdd12.pdf>.

[3] "Adamic et al., Knowledge Sharing and Yahoo Answers: Everyone Knows Something, WWW 2008 <http://snap.stanford.edu/class/cs224w-readings/adamic08knowledge.pdf >

[4] Morris et al., What Do People Ask Their Social Networks, and Why? A Survey Study of Status Message Q&A Behavior, SIGCHI Conference on Human Factors in Computing Systems <http://research.microsoft.com/en-us/um/people/teevan/publications/papers/chi10-social.pdf >

[5] Jurgens et al., An Analysis of Exercising Behavior in Online Populations, AAAI Conference on Web and Social Media, 2015 <http://www-cs.stanford.edu/jurgens/docs/jurgens-mccorriston-ruths_icwsm-2015.pdf >

[6] Cunha et al., The Effect of Social Feedback in a Reddit Weight Loss Community, 6th International Conference on Digital Health, 2016 <https://arxiv.org/pdf/1602.07936.pdf >

[7] Althoff et al., Online Actions with Offline Impact: How Online Social Networks Influence Online and Offline User Behavior WSDM 2017

[8] Tang et al., Building Large-Scale Twitter-Specific Sentiment Lexicon : A Representation Learning Approach, 25th International Conference on Computational Linguistics 2014

[9] Minqing Hu and Bing Liu. "Mining and Summarizing Customer Reviews." Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD-2004), Aug 22-25, 2004, Seattle, Washington, USA,

[10] David M. Blei and Andrew Y. Ng and Michael I. Jordan. "Latent Dirichlet Allocation." The Journal of Machine Learning Research. Pages 993-1022. Volume 3, 3/1/2003.

[11] Support-vector networks C. Cortes and V. Vapnik. "Support-vector networks". Machine learning, 1995.

[12] A Clauset and MEJ Newman and C Moore. "Finding community structure in very large networks." Physical review E 70 (6), 066111.

[13] Ader et al., "Why We Search: Visualizing and Predicting User Behavior." WWW 2007.