2013 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining

Game-Theoretic Approach for User Migration in Diaspora

Mohammad Rashedul Hasan, Mohamed Shehab, Ali Noorollahiravari
Department of Software and Information Systems
University of North Carolina at Charlotte
Charlotte, NC 28233, USA
Email: {mhasan8, mshehab, anooroll}@uncc.edu

***Abstract*—Diaspora is a decentralized online social networking platform where user profiles are hosted in multiple Diaspora nodes (pods) and the social connections can exist across different pods. User profile migration is a promising feature that would enable users to seamlessly migrate their profile data between different pods. However, to the best of our knowledge, there has been no research done on how this data portability may affect the user distribution and the performance of the pods. In this paper, our goal is to design an approach that facilitates the users to choose appropriate pods that would ensure better service quality. We propose a decentralized game-theoretic approach that is based on user's local neighborhood information and the quality of the pods. We have analytically determined, and experimentally substantiated, that through the proposed profile migration approach the users of Diaspora reach a stable and balanced distribution that improves their overall experience in respective pods.**

***Index Terms*—Diaspora, Online Social Network, Game Theory, Data Portability.**

I. Introduction

In order to address the concern for the growing potential of privacy and security breaches in centralized online social networks (OSNs), several decentralized OSNs or DOSN architectures have been proposed. Among them Diaspora [1], [2] is the only large-scale DOSN that currently has approximately 400,000 active users. It is a privacy aware open source DOSN that promises its users to have control over their data. In Diaspora, the user data could be under complete control of the user or it could be administered by multiple administrative domains. Although Diaspora has emerged as a promising privacy-preserving alternative, there is little research on its performance and architecture related issues that need to be addressed. For example, in Diaspora network when a user shares any post with its intended recipients who belong to different servers (also referred to as pods), the user's hosting pod has to send that post to those pods. After that any comment made on that post by any recipients has to be replicated and distributed to all the original post recipients. However, so far no investigation has been done to assess the impact of this distributed data replication over multiple pods. Also, it's not clear how this type of user activity could exhaust the pod resources and whether it's possible to derive any decentralized solution.

Recently Bielenberg et al. [3] conducted an empirical study on the Diaspora network and observed that its user distribution is highly unbalanced. For example, 70% of the users reside on the main server developed by the Diaspora providers (http://joindiaspora.com) and approximately 94% of the total Diaspora users reside in the four largest servers of this DOSN. In other words, a small percentage of users are hosted by majority of the servers, while the majority of the users are hosted by very few large servers. As a consequence, larger servers (with more than 100 users) have a smaller share of replicated profiles while in smaller servers the number of replicated profiles are too high. Because of this unbalanced user distribution there would be a large amount of data replication that may exhaust the pod resources.

As a solution to this problem, a user may migrate to an appropriate pod that would reduce the inter-pod data traffic without sacrificing the quality of service. For example, by moving the user's profile to a pod containing most of his friends will effectively reduce the inter-pod traffic. However, the current implementation of the Diaspora does not allow the users to seamlessly migrate their data to the pod of their choice and promise to offer this service in future.

In this paper, we address this issue of user profile migration between different pods in Diaspora and propose a migration algorithm that enables the users to move to a pod that would improve their overall experience. We use a decentralized game-theoretic approach for the users to evaluate the payoff of each pod with respect to the various groups in their profile. The payoff is defined by the users' neighborhood information and the quality of each pod. For example, if the majority of the friends of a user reside in a different pod, the user may prefer to move to that pod. However, this may lead to an unbalanced user distribution of the network favoring few pods over the others; and due to the large concentration of the users, those few pods would suffer performance degradation. Therefore, it is important to make a trade-off between a preferred pod and its quality-of-service. Using the proposed game-theoretic approach we intend to determine an optimum trade-off that not only ensures better user distribution, but also retains better pod performance.

Recently decentralized game-theoretic techniques have been used to solve the community detection or clustering problem in large networks including the OSN [4], [5]. The nodes are considered as independently functioning rational entities trying to optimize their utility/payoff by joining/leaving a community. These approaches use only local information

ASONAM'13, August 25-29, 2013, Niagara, Ontario, CAN

about the neighborhood of the nodes. We adopt a similar approach where the users of the DOSN are considered as self-interested and rational players. We design a Pod Selection Game (PSG) in which the users play the game to optimize their payoff. We show that it is possible to reach a Nash Equilibrium distribution of the users that increases the total payoff of the Diaspora network. Our main contributions are following:

- We identified the user profile migration problem in distributed online social networks.
- We modeled and formulated the Pod Selection Game (PSG) that guides users to select pods that reduces amount of data replication.
- We have shown analytically, and substantiated experimentally, that the PSG converges into a Nash Equlibirium distribution.

The remainder of this paper is organized as follows. Section IV presents our proposed decentralized Pod Selection Game algorithm and an analytical discussion on the Nash Equilibrium user distribution. In Section V, we provide the simulation results and discuss the effectiveness of the proposed game. We conclude the paper in Section VI by summarizing our contributions and identifying future challenges.

II. Preliminaries

Figure 1 shows a basic overview of the Diaspora pods and their communication. Each pod can be considered as a small social network of trusted users. The users' social network can extend across several pods. Each pod is able to communicate with the users in other pods. The communications in Diaspora are conducted through posts. The privacy of a post could be either public or private. A public post can be accessed by any user in the network, while a private post has a specific group of audience.

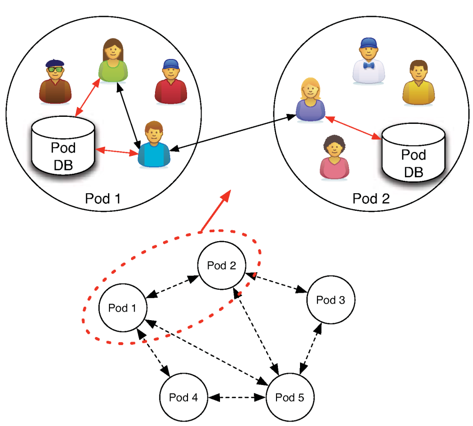


Fig. 1. Diaspora pods and their communication

A push design mechanism is used to distribute data and notifications to other pods in the network. The pod where the original post is created will be responsible for the replication and relaying of the comments related to that post. The post is shared with users by pushing out the content to all of the servers hosting the intended users. In this way, diaspora shares duplicates of the same data on certain different pods. As a consequence, the network is confronted with the following issues: (a) the security and integrity of a post is dependent on the administrators of the servers where it is stored, (b) if most of the members are situated in different pods, then this would require a large amount of replication to push data to all the target pods and (c) pods are required to provide resources to manage push notifications to both local and foreign users.

III. Related Works

The issue of characterization and detection of community structures in social networks has received a significant amount of attention in the literature [6]. The classic problem in social networks is to divide the nodes of the network into some number of groups, and to minimize the number of interconnections between them. In this way, a densely connected family of vertices is detected to be in each group. Game-theoretic techniques has been successfully used to solve this community detection problem in social networks [7], [4], [5]. In [5] a community detection game is set up for each component in the network that depends on the richness of the neighborhood of each node. In this way, the quality metric is completely dependent on the local information about the neighborhood of each component. [7] introduced a game-theoretic framework to address the problem based on the structures of social networks. By using a simple utility function they show that their method is effective in detecting overlapping communities. In addition, the algorithm operates on local data to reach equilibrium, and its running time is fast. [4] proposes a new game theoretic framework for the community detection problem. The focus here is on the convergence of fair and stable payoff functions instead of the optimization of a global function. Through experimental results they show that their algorithm ensures strict modularity optimization.

Profile migration of users across different pods in a DOSN can be modeled as a clustering problem in social networks, where the pods are the clusters and the users need to decide which cluster to migrate to. Our proposed profile migration approach is inspired from the clustering models in social networks but our objective is not only to optimize the network structure but also to investigate how the user distribution would affect the pod efficiency.

IV. Proposed Approach

In this section we present our proposed game theoretic approach to facilitate user profile migration across pods, which we refer to as the *Pod Selection Game (PSG)*.

A. The Model

The users of the different Diaspora pods are considered as the players in the Pod Selection Game. The players find and move to the pod that not only reduces the amount of data replication and improves her/his privacy, but also improves the users' overall experience in that pod. Our approach is based on the user's neighborhood information (friend list) and the pods' quality and efficiency. We anticipate that a user may want to move a pod where most of her/his friends reside. However, instead of considering the number of neighbors we use a fine-grained approach to find cohesive subgroups (referred to as **aspects**), which are sharing groups similar to Google+ groups whose members interact more frequently with each other than with those outside the group. In other words, we consider that a user may want to switch to a pod where most of the members of her/his preferred aspect belongs (that has more intra-links among the members of that aspect). But this pod selection strategy may create an unbalanced user distribution and degrade the performance of the pods where most of the users has moved. That's why apart from using the neighborhood information we also use the quality of the pods as a determining criterion for the user. This is captured through the node payoff function that we define in the following.

We consider an undirected and unweighted network $G =< N, E >$, where N is the finite set of nodes or players and E is the set of edges. Players are organized into two types of clusters, namely pods and aspects. We assume that the set of pods is $P = \{1, 2, 3, \cdots, p\}$ and the set of aspects is $A = \{1, 2, 3, \cdots, a\}$ existing in G. The pair (p_j, a_k), where $j \in P$ and $k \in A$, denotes a player's neighborhood information. For example, a players adjacency list could be distributed among multiple pods and multiple aspects.

For each player a real-valued payoff function u_i is used to determine the player's payoff for every possible neighborhood in G. We define the payoff function as $u_i : (p_j, a_k) \to R$, where $i \in N$, $j \in P$ and $k \in A$. The pod selection game should allow the players to optimize their payoff by moving to a preferred pod and should take into consider the pod quality.

The payoff function is defined as the difference between the benefit and the cost [8]. The benefit portion is captured through the normalized number of pairs of the neighbors belonging to an aspect inside a pod. The cost portion is captured through a weight function that represents the pod's quality [8]. For each player $i \in N$ belonging to the aspect a_k and pod p_j, the payoff is defined as following: $u_i = \alpha \frac{F(p_j, a_k)}{G(p_j, a_k)} - \beta T(p_j)$

where $F(p_j, a_k)$ represents the number of pairs of the neighbors of the player i that belongs to the aspect a_k and pod p_j and $G(p_j, a_k)$ represents the number of every possible pair. The intuition is that the user benefits by moving to pods that host most of his friends as this will require less replication and will provide better experience. The parameters α and β are scaling factors for the benefit and cost functions, respectively. By varying these two parameters a trade-off can be made between the choice of preferred neighborhood and a pod's quality. $T(p_j)$ is the weight function for pod p_j and is defined as following: $T(p_j) = \exp(\gamma_1 \frac{M(p_j)}{M(p_j)+N(p_j)} + \gamma_2 \frac{X(p_j)}{X(p_j)+Y(p_j)})$

where $M(p_j)$ is the number of Home users in pod p_j, $N(p_j)$ in the number of Foreign users in p_j, $X(p_j)$ is the number of outgoing links from p_j and $Y(p_j)$ is the number of intra-pod links in p_j. The scaling parameters γ_1 and γ_2 control the effect of the number of pod users and cost of replication respectively. The intuition behind this model is that the pod cost increases when the number of users it services increase and when the number of outgoing links (data replications) increases. Note that the parameters used to compute both the benefit and cost can be calculated based on the local network neighborhood information of the user and the pod's efficiency metrics (such as the number of home and foreign users; and the outgoing links). We assume that the pod's efficiency metrics are advertised to the users.

Given the payoff function u_i, at each stage of the game the users optimize their payoff by migrating to preferred pods.

B. Equilibrium and Algorithm

By showing that there exists a Nash Equilibrium distribution we can conclude that the Pod Selection Game ensures the convergence into a stable node distribution over the set of pods. In other words, we prove that this game has a Nash Equilibrium solution. We define the Nash Equilibrium solution as following:

Definition 1. *Nash Equilibrium: A pod P_j is a Nash Equilibrium if all nodes of this pod that has membership in group A_k satisfies the following condition. For convenience we designate by subscript $j(-k)$ all the other pods (aspects) in the network.*

$$u_i(p_j, a_k) \geq u_i(p_{-j}, a_k), \forall i \in p_j$$

Proposition 1. *The Pod Selection Game converges into a Nash Equilibrium node distribution.*

Proof: Let $\phi(P_j) = \sum_{i \in P_j, k \in P_j(Aspects)} u_i(A_k)$ be the sum of the payoffs of all the aspects of every node belonging to a pod p_j. Let us consider two pods P_1 and P_2. Now whenever, based on the PSG, a node i moves from P_1 to P_2, its payoff should increase as $u_i(p_2, a_k) > u_i(p_1, a_k)$

$\implies \alpha \frac{F(p_2, a_k)}{G(p_2, a_k)} - \beta T(p_2) > \alpha \frac{F(p_1, a_k)}{G(p_1, a_k)} - \beta T(p_1)$

This is possible when $F(p_2, a_k) > F(p_1, a_k)$ or $T(p_2) < T(p_1)$. A node's payoff increases if it moves to a pod where the neighbors of its aspect(s) are mostly connected. However, if this migration degrades the pod's service quality (with increasing number of users and outgoing links), the cost associated with the payoff increases. In that case, the node would try to minimize the cost by migrating to a pod with smaller pod weight function value. Therefore, node migration would always tend towards increased payoff. This ensures that the sum of the payoffs of nodes moved into a pod is greater than the sum of the payoff of their previous pods.

As a consequence $\phi(P_2) > \phi(P_1)$. Therefore, one node distribution clearly dominates the other distribution. Since there are finite number of node distributions, there exists a Nash Equilibrium distribution.

■

In the above proof, we assumed that the values of the parameters α, β, γ_1 and γ_2 are set to 1.0. However, variation of these parameters could result in different user distributions. The experiments on various synthetic networks that we report in the next section indicates that when $\alpha > \beta$, the quality of the pods play a significant role for choosing an appropriate pod. With $\alpha = \beta = 1.0$, the network converges into a skewed user distribution favoring one or few pods over the others. In this case users put more emphasis on the neighborhood information than the quality of the pods. Therefore, it appears that there could be multiple Nash Equilibrium distributions and the values of α and β could be used to determine the preferred distribution. For our experiments we found better results by setting $\alpha = 5.0$ and $\beta = 1.0$.

V. Experiments and Evaluation

In order to perform the simulation on the nodes of the Diaspora network we require every user or node to have the adjacency list of all the nodes. In other words, we require the nodes to have the friend lists of all the nodes. However, the privacy-preserving design policy of the Diaspora network does not allow us to see the friend list of the users. That is why we were not able to obtain the link information between the users of the real Diaspora network. Therefore, we decided to run our simulation on synthetic networks that resemble the major characteristics of the Diaspora network. We investigated several types of networks ranging from a balanced node distribution, where the user nodes are uniformly distributed over the available pods, to an unbalanced node distribution where the majority of the nodes are concentrated in few pods, as in the case of current user-distribution in the Diaspora network [3].

The link structure or the degree-distribution in the existing OSNs (e.g., Facebook) follows power-law form [9]. Therefore, we create a power-law scale-free network by using the Barabasi-Albert model [10]. The number of nodes in our synthetic network is fixed to 5000. We create six pods, namely A, B, C, D, E and F; and distribute the nodes into these pods. We distribute the neighbors of the nodes uniformly into three aspects, namely Family, Friends and Work. The pod and aspect memberships are randomly assigned based on a uniform distribution. Each simulation consists of 100 time steps where a time step refers to a single run of the program. The following values are assigned to the parameters of the payoff function: $\alpha = 5.0$, $\beta = 1.0$, $\gamma_1 = 1.0$ and $\gamma_2 = 1.0$.

Balanced Node Distribution. We consider two scenarios for balanced node distribution: (a) in Network 1 all the nodes are uniformly distributed into six pods and (b) in Network 2 all the nodes are uniformly distributed into five pods while the sixth pod contains only one node.

Unbalanced Node Distribution. A recent study on the existing Diaspora network shows that more than 70% users are situated in one pod and that the largest four servers contain 94% of the users [3]. In this node distribution, there are a small number of pods that contain most of the network users. In order to simulate such unbalanced user distribution we consider two scenarios: (a) in Network 3, we assign 70% of the nodes into a single pod A and (b) in Network 4, 96% of the nodes are uniformly distributed among two pods (A and B).

Network 1: Balanced Node Distribution, Figure 2(a) shows the initial and final node distributions for Network 1. The balance in the initial node distribution does not appear to change much in the final state. Figure 2(d) shows the number of users in each pod after executing a PSG step. We notice that within 10 time-steps the network converges into a stable state.

Network 2: Balanced Node Distribution, Figure 2(b) shows the initial and final node distributions for Network 2. In the initial node distribution, pod F contains only 1 user node. In this network all the pods have similar number of users except pod F. Here we investigate a scenario in which a new pod (pod F with only one user) is added to the Diaspora network. After performing the PSG steps as the network reaches a converged state, pod F gets 355 nodes and the overall node distribution improves significantly. Figure 2(e) shows the number of users in each pod after executing a PSG step. We notice that within 10 time-steps the network converges into a stable state.

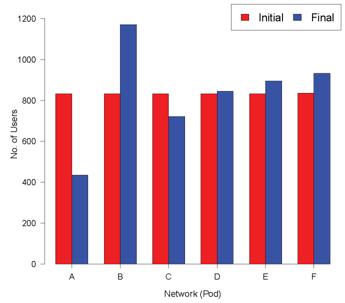
Network 3: Unbalanced Node Distribution Figure 2(c) shows the initial and final node distributions for Network 3. This network is unbalanced because most (70%) of the user nodes are hosted in pod A. After performing the PSG steps and reaching the convergence state we notice that the number of users in pod A is reduced and these users are distributed among the other pods. Figure 2(f) shows the number of users in each pod after executing a PSG step. The network converges into a stable state within 10 time-steps.

Network 4: Unbalanced Node Distribution Figure 2(g) shows the initial and final node distributions for Network 4. This network is unbalanced because most (96%) of the user nodes are hosted in pods A and B. After performing the PSG steps and reaching the convergence state we notice that the number of users in pods A and B are reduced and these users are distributed in other pods. Note that the number of users in pod A and pod B are not reduced by the same factor; we attribute this to the difference in the network structure and the number of inter and intra-pod links of these pods. Figure 2(h) shows the user distribution among the pods after executing a PSG step. The network converges into a stable state within 10 time-steps.

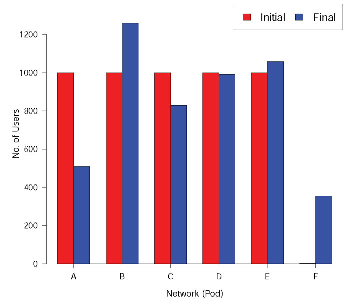
Discussion: The experiments are based on synthetic networks and are not a comprehensive experimentation of all possible networks. However, we tried to model the real world Diaspora network scenarios based on the available statistics. Figure 2(i) shows how the total payoff across various networks changes after reaching a converged state. In Network 1, the increase in the total payoff is not high, because the initial node distribution was uniform. On the other hand, we notice a significant increase in the total payoff in Network 3, because the initial user distribution was highly unbalanced. In Network 2, initially one pod had only one user, that's why the initial total payoff of the network was low. However, after playing the PSG, we notice a significant increase in the total payoff as the network assumed a balanced distribution.

VI. Conclusion and Future Work

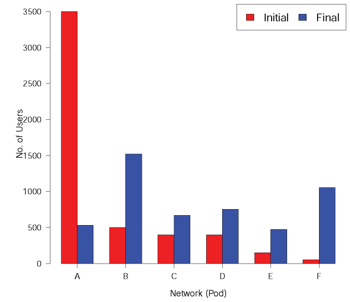
In this paper, we designed a decentralized game-theoretic approach to facilitate the users of the DOSN Diaspora to migrate to an appropriate pod. We have proposed a Pod



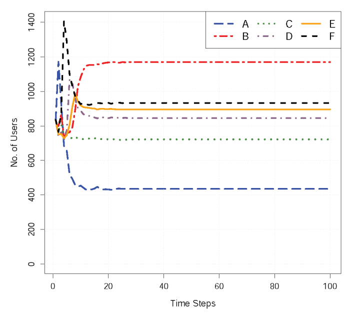
(a) Net 1 Distribution



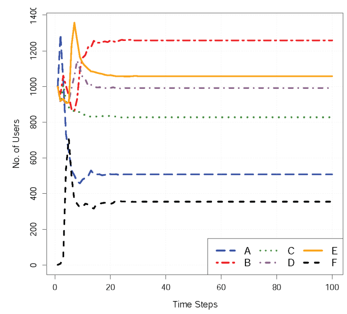
(b) Net 2 Distribution



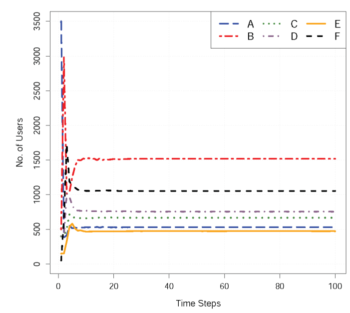
(c) Net 3 Distribution



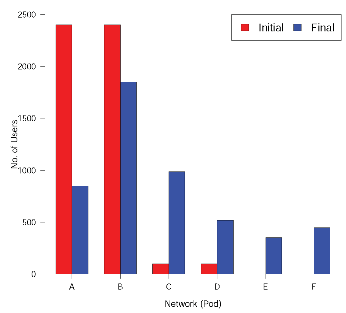
(d) Net 1 Convergence



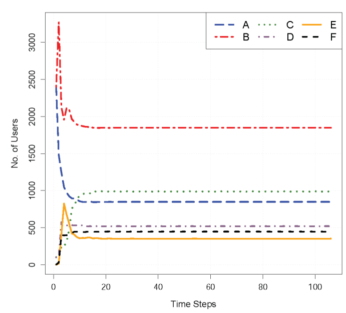
(e) Net 2 Convergence



(f) Net 3 Convergence



(g) Net 4 Distribution



(h) Net 4 Convergence

Network	Δ = Final Total Payoff - Initial Total Payoff
Network 1	119.26
Network 2	1067.38
Network 3	17679.58
Network 4	12632.03

(i) Change in total payoff after PSG

Fig. 2. Networks 1, 2, 3 and 4 initial/final distributions and convergence

Selection Game (PSG) based on the user's local neighborhood information and pod's quality metrics in which self-interested and rational users intend to optimize their payoff. We have defined the payoff metric as a cost-benefit function in which migrating to the pod of a dense aspect might increase the benefit but the consequential accumulation of large number of users in few pods increases the cost by degrading user experience. Analytically we proved that the PSG ensures a Nash Equilibrium user distribution. Using various types of synthetic networks we have experimentally substantiated this claim by showing that such type of user distribution increases the total payoff of the pods. As future work, we intend to design a *group migration* framework where instead of migrating individually an entire group of user may benefit more by migrating in groups. Also, we will investigate user swapping among pods.

VII. Acknowledgements

This research was partially supported by grants from the National Science Foundation (NSF-CNS-0831360, NSF-CNS-1117411) and a Google Research Award.

References

[1] (2013) Diaspora. [Online]. Available: http://diasporaproject.org/
[2] A. Bleicher, "The anti facebook," *IEEE Spectrum*, vol. 48, no. 6, pp. 54–82, June 2011.
[3] A. Bielenberg, L. Helm, A. Gentilucci, D. Stefanescu, and H. Zhang, "The growth of diaspora - a decentralized online social network in the wild." in *INFOCOM Workshops*. IEEE, 2012, pp. 13–18.
[4] P. J. McSweeney, K. Mehrotra, and J. C. Oh, "A game theoretic framework for community detection." in *IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM 2012)*. IEEE Computer Society, August 2012, pp. 227–234.
[5] R. Narayanam and Y. Narahari, "A game theory inspired, decentralized, local information based algorithm for community detection in social graphs," in *ICPR*, 2012, pp. 1072–1075.
[6] M. K. Goldberg, S. Kelley, M. Magdon-Ismail, K. Mertsalov, and A. Wallace, "Finding overlapping communities in social networks," in *SocialCom/PASSAT*, 2010, pp. 104–113.
[7] W. Chen, Z. Liu, X. Sun, and Y. Wang, "A game-theoretic framework to identify overlapping communities in social networks," *Data Min. Knowl. Discov.*, vol. 21, no. 2, pp. 224–240, Sep. 2010.
[8] J. Freudiger, M. H. Manshaei, J.-P. Hubaux, and D. C. Parkes, "On non-cooperative location privacy: a game-theoretic analysis," in *ACM Conference on Computer and Communications Security*, 2009.
[9] S. Catanese, P. D. Meo, E. Ferrara, G. Fiumara, and A. Provetti, "Extraction and analysis of facebook friendship relations," 2011.
[10] A.-L. Barabási and R. Albert, "Emergence of scaling in random networks," *Science*, vol. 286, pp. 509–512, 1999.