Self-Organizing Flows in Social Networks

Nidhi Hegde, Laurent Massoulié, Laurent Viennot

Self-Organizing Flows in Social Networks*

Nidhi Hegde
Technicolor
Paris, France
nidhi.hegde@technicolor.com

Laurent Massoulié†
Microsoft Research – Inria Joint Centre
France
laurent.massoulie@inria.fr

Laurent Viennot‡
Inria – Paris Diderot University
France
laurent.viennot@inria.fr

Abstract

Social networks offer users new means of accessing information, essentially relying on "social filtering", i.e. propagation and filtering of information by social contacts. The sheer amount of data flowing in these networks, combined with the limited budget of attention of each user, makes it difficult to ensure that social filtering brings relevant content to interested users. Our motivation in this paper is to measure to what extent self-organization of a social network results in efficient social filtering.

To this end we introduce *flow games*, a simple abstraction that models network formation under selfish user dynamics, featuring user-specific interests and budget of attention. In the context of homogeneous user interests, we show that selfish dynamics converge to a stable network structure (namely a pure Nash equilibrium) with close-to-optimal information dissemination. We show that, in contrast, for the more realistic case of heterogeneous interests, selfish dynamics may lead to information dissemination that can be arbitrarily inefficient, as captured by an unbounded "price of anarchy".

Nevertheless the situation differs when user interests exhibit a particular structure, captured by a metric space with low doubling dimension. In that

*Supported by PROSE ANR project.
†Part of this work was done while at Technicolor.
‡Supported by the Inria project-team "Gang" and by the joint laboratory "LINCS".

case, natural autonomous dynamics converge to a stable configuration. Moreover, users obtain all the information of interest to them in the corresponding dissemination, provided their budget of attention is logarithmic in the size of their interest set.

Keywords: Network formation, self organisation, budget of attention, price of anarchy, social filtering.

1 Introduction

Information access has been revolutionized by the advent of social networks such as Facebook, Google+ and Twitter. These platforms have brought about the new paradigm of "social filtering", whereby one accesses information by "following" social contacts.

This is especially true for twitter-like microblogging social networks. In such networks the functions of filtering, editing and disseminating news are totally distributed, in contrast to traditional news channels. The efficiency of social filtering is critically affected by the network topology, as captured by the contact-follower relationships. Today's networks provide recommendations to users for potentially useful contacts to follow, but don't interfere any further with topology formation. In this sense, these networks self-organize, under the selfish decisions of individual users.

This begs the following question: when does such autonomous and selfish self-organizing topology lead to efficient information dissemination? The answer will in turn indicate under what circumstances self-organization is insufficient, and thus when additional mechanisms, such as incentive schemes, should be introduced.

Two parameters play a key role in this problem. On the one hand each user aims to maximize the coverage of the topics of his interest. On the other hand, a user pays with his attention: filtering interesting information from spam (i.e. information that does not fall in his topics of interest) incurs a cost. Users must therefore trade-off topic coverage against attention cost. As pointed out by Simon [22], as information becomes abundant, another resource becomes scarce: attention.

Furthermore, there is an interplay between participants in a social network where filtering by one user may benefit another, inducing complex dependencies in decisions on creating connections. To model this, we introduce a network formation game called *flow game* where some users produce news about specific topics and each user is interested in receiving all news about a set of topics specific to him. Each user is a selfish agent that can choose its incoming connections within a certain budget of attention in order to maximize the coverage of his set of topics of interest.

This model is of interest on its own, as it enriches the class of existing network formation games with a focus on flow dissemination under bounded connections. This model could also be of interest in the context of peer-to-peer streaming and file sharing or publish/subscribe applications.

1.1 Our results

An important feature in our model is a user's budget of attention for the consumption of content. In previous work [13] the budget of attention was modelled as a limit on the rate with which a user consults a friend, with a different objective of minimizing delay in receiving all content. In the present work we are interested in a more fundamental question, of how efficient social networks are formed in the first place. We consider the model where users are interested in specific subsets of topics and their objective is to maximize the number of flows received corresponding to these topics. As such, we model the budget of attention as a constraint on the number of connections a user may create (rather than a rate of consultation). Our aim is to build a simple model capturing the complexity of the problem. This way of capturing the budget of attention amounts to assuming that each connection consumes the same amount of attention. We discuss in the conclusion how we could tweak our model to more finely model attention consumption.

We capture users' interests in topics through user-specific values for each topic and define the *utility* a user receives to be the sum of values of all received topics. Each user's objective in a *flow game* is then to choose connections so as to maximize his utility. We additionally assume that a user may produce news about one topic at most even if he redistributes other topics. This is coherent with an empirical study of twitter traces [5] where it is shown that ordinary users (as opposed to celebrities or newspapers) can gain influence by concentrating on a single topic.

Our main results relate to the stability and efficiency of the formation of information flows. We derive conditions where selfish dynamics converge to a pure Nash equilibrium.We then give approximation ratios bounding the quality of an equilibrium compared to an optimal solution. This is traditionally measured through the price of anarchy, the ratio of the global welfare (measured as the sum of user utilities) at an optimal solution compared to that at the worst equilibrium.

1.2 Related work

Information spread in networks has been studied extensively. Much of the past work study the properties of information diffusion on given networks with given sharing protocols. Our goal in this work is to study how networks form when users create connections with the objective of efficient content dissemination in a game-theoretical approach. This work thus follows the large amount of work in network formation games. However, to the best of our knowledge, the objective of efficient information dissemination under edge constraints and interest sets that we consider here is novel. We now discuss some work in those domains that are most relevant to this paper.

Network formation games have been considered in previous work in economics and in the context of the formation of Internet peering relations and peer-to-peer overlay networks. Economic models of network formation [12] use edges to represent social relations and it is typically assumed that the creation of an edge needs bilateral agreement since both users benefit from an edge. Our model is oriented and unilateral agreement is more relevant to the notion of *following* in social networks. A non-cooperative one-way link connection game has been considered been in previous work [3], where each created link incurs a cost and users are interested in connecting to all other users. Our model is richer and more realistic where we consider connections to subsets of information flows that hold user-specific intrinsic values.

Network creation games in the context of the Internet have been considered [18], where distributed formation of undirected edges with a linear cost on each edge formed is studied. In such games, each user's objective is to minimize total formation cost while either minimizing distance to all other users [7], or ensuring connection to a given subset of nodes [2]. We consider a bound on edge costs, in the form of a limit on the number of in-edges at each node, and further, we focus on connections that allow specific flows of information.

Interestingly, bounded budget network formation games have already been considered. Bounded budget connection games [15] consider a bound on each user's budget in creating edges, with the objective being the minimization of the sum of weighted distances to other nodes. A similar model is considered in [4] where each user's objective is to maximize his influence, measured using betweenness centrality. In our work however, rather than minimizing distance to any node, we consider a formation game with the objective of ensuring connections to a subset of flows of interest, without regard to the particular nodes.

The notion of connecting to users that can provide a content flow of interest is similar to peer-to-peer live streaming systems [16]. Unlike peer-to-peer streaming, we do not aim to satisfy flow rates, rather our aim is to connect to as many sets of relevant flows as possible. Moreover, our model allows differing user interests.

To the best of our knowledge the only work considering content dissemination with some game-theoretical approach concerns the b-matching and acyclic preference systems studied in the context of peer-to-peer applications [10]. As a generalization of the stable marriages problem, those systems consider configurations of undirected edges based on mutual acceptance of an edge, whereas unilateral decision is more suitable in our model. Our model is more intricate in the sense that connections are based not only on preferences but also on complementarity of content obtained through various connections.

In Section 5 we model the space of user interests by a metric space with low doubling dimension. Modeling interests of users through a metric space seems a natural approach and bounded growth metrics, or more generally doubling metrics, have shown to be very a general model [19] that can capture general situations, while still providing an algorithmic perspective. The doubling dimension extends the notion of dimension from Euclidean spaces to arbitrary metric spaces. It has proven to be useful in many application domains such as nearest neighbor queries to databases [6], network construction [1], closest server selction [14], etc. Doubling metrics have notably been used to model distances in networks such as Internet [9].

1.3 Organization of the paper

Section 2 introduces the model. We study the case of homogeneous interests in Section 3. The heterogeneous case in its full generality is considered in Section 4 which details some negative results. Section 5 is dedicated to the specific scenario where users' interests are captured by a doubling metric, enabling some positive results. We finally conclude in Section 6 describing potential extensions of the current work.

2 Model

We consider a social network where users interested in some set of content topics (or subjects) connect to (or *follow* in social networking parlance) other users in order to obtain such contents, materialized by flows of news. Each user may produce news for at most one topic (but may forward news from other topics she is interested in). To distinguish the role of publisher from that of follower, we technically assume that news concerning a given topic (or subject) are produced at a given node called producer which is identified with that topic.

A *flow* game is defined as a tuple (V, P, S, Δ) where V is a set of users, P a set of producers (or subjects or topics) and $S : V \to P$ is a function associating to each user u its interest set $S_u \subseteq P$, and $\Delta : V \to \mathbb{N}$ is a function associating to each user u its budget of attention Δ_u. We let $n = |V|$ and $p = |P|$ denote the number of users and producers respectively. A flow game is *homogeneous* if all users have the same interest set: $S_u = P$ for all $u \in V$. If this is not the case, the game is said to be *heterogeneous*.

A strategy for user u is a subset F_u of $\{(v, u) : v \in V \cup P\}$ such that $|F_u| \leq \Delta_u$ (Δ_u is an upper bound on the in-degree of u). For all $(v, u) \in F_u$, we say that u *follows* v or equivalently that u is connected to v (such a link (v, u) created by u is oriented according to the data flow, that is from v to u). The collection $F = \{F_u : u \in V\}$ forms a network defined by the directed graph $G(F) = (V \cup P, E(F))$ where $E(F) = \cup_{u \in V} F_u$. A user u is *interested* in a subject s if $s \in S_u$. A user u *receives* a subject $s \in P$ if there exists a directed path from s to u in $G(F)$ such that all intermediate nodes are interested in s. This is where *filtering* occurs: a user retransmit only subjects she is interested in. The utility $U_u(F)$ for user u is the number of subjects in S_u she receives. The utility of u is maximized if $U_u(F) = |S_u|$.

We denote by *move*, a shift from a collection F of strategies to a collection F' where a single user u changes her strategy from a set F_u to another F'_u. (We say that u rewires her connections.) The move is *selfish* if $U_u(F') > U_u(F)$. *Selfish dynamics* (or dynamics for short) are the sequences of selfish moves. We say that dynamics *converge* if any sequence of selfish moves is necessarily finite. The network is at equilibrium (or stable) if no selfish move is possible. In standard game-theoretic terminology, this corresponds to a pure Nash equilibrium. The *global welfare* of the system is defined as the overall system utility: $\mathcal{U} = \sum_{u \in V} U_u$. The efficiency of selfish, self-organization of a game is classically captured by the notion of price of anarchy defined as the ratio of the optimal global welfare over the global welfare of the worst equilibrium: $\text{PoA} = \frac{\max_{F \in \mathcal{F}} \sum_{u \in V} U_u(F)}{\min_{F \in \mathcal{E}} \sum_{u \in V} U_u(F)}$, where $\mathcal{F}$ denotes the set of possible collection of strategies and $\mathcal{E} \subseteq \mathcal{F}$ denotes the set of equilibria.

In some of our proofs we make use of the notion of potential functions. An ordinal (or general [8]) potential function [17] is a function $f : \mathcal{F} \to \mathbb{R}$ such that $sign(f(F') - f(F)) = sign(U_u(F') - U_u(F))$ for any move from F to F' where user u changes her strategy. If $f(F') - f(F) = U_u(F') - U_u(F)$, f is called an exact potential function. This notion was introduced by Monderer and Shapley [17] who show that it is tightly related to the notion of a congestion game [20]. The use of potential functions is a standard technique to show convergence of dynamics and to bound price of anarchy [8, 21].

3 Homogeneous interests

We first consider the case where all users have identical sets of interests, $S_u = P$, for all $u \in V(G)$. In this context, we first establish an upper bound on the price of anarchy. We will then show convergence of dynamics.

3.1 Price of Anarchy

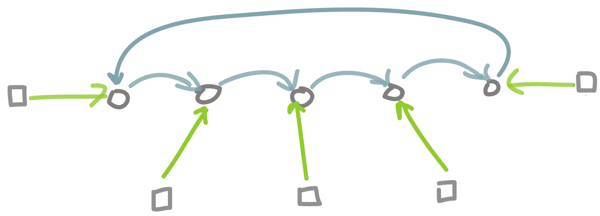
We first derive a simple upper bound on the overall system utility under an optimal centrally designed configuration. Clearly, any user u cannot achieve utility larger than p, which corresponds to obtaining all the subjects in P. Moreover, he cannot obtain more subjects than the aggregate budget of attention of all users, that is $\sum_{u \in V(G)} \Delta_u = n\overline{\Delta}$, where $\overline{\Delta}$ is the average in-degree per node. We can slightly improve this bound by restricting ourselves to the more interesting case where all users have budget less than p and where there are at least two users with budget at least 2. One can easily see that the optimal solution then consists in forming an oriented ring between users whose budget is at least 2 and then connecting budget 1 users to some user of the ring. All remaining connections are used to obtain distinct subjects. Each node then receives the same set of subjects. As each node connects to a non-producer, the number of subjects gathered is at most $\sum_{u \in V(G)} \Delta_u - 1$. We thus obtain that the maximal utility U^* a user can get is:

$$U^* = \min\left(p, n(\overline{\Delta} - 1)\right) . \quad (1)$$

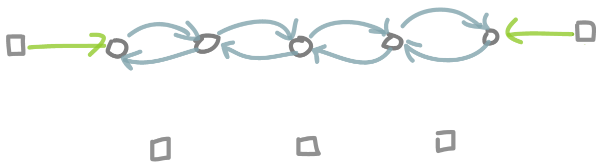
We now consider a distributed setting where each user selfishly rewires his incoming connections if he can improve his utility, i.e., if this allows him to receive more subjects. The following proposition shows that with homogeneous user interests and budget of attention at least 3, self organization is efficient if dynamics converge, achieving a price of anarchy close to 1.

Proposition 1 *Assume that $3 \leq \Delta_u < p$ for every user $u \in V$ of a homogeneous flow game. Then under any equilibrium the utility of a user is at least $\frac{\overline{\Delta}-2}{\overline{\Delta}-1} U^*$ where U^* is his optimal utility. The price of anarchy is thus at most $1 + 1/(\overline{\Delta} - 2)$, approaching 1 for large $\overline{\Delta}$.*

We first note that the above proposition is tight in the sense that high price of anarchy can arise when $\Delta_u \leq 2$ for all user u, as shown in Figure 1. In this particular case, a doubly linked chain forms a Nash equilibrium gathering only two subjects in total while an oriented cycle gathers n subjects. The price of anarchy is thus $n/2$.



(a) Benchmark configuration



(b) A Nash equilibrium configuration

Figure 1: Homogeneous interest sets with degree $\Delta = 2$.

Before proving Proposition 1, we establish two lemmas. To establish Proposition 1, we use two lemmas. The first one allows to show the existence of strongly connected components at equilibrium.

Lemma 1 *If an equilibrium is reached such that there exists a path $x, u_1, \dots, u_k$ where x is a producer, u_k has in-degree bound $\Delta_{u_k} \geq 3$ and a producer y is not received by u_k, then there is a path from u_k to u_1.*

Proof. The existence of the path $x, u_1, \dots, u_k$ first implies that $R(u_1) \subset R(u_k)$. Since $\Delta_{u_k} \geq 3$, u_k must be connected to two nodes v and w distinct from u_{k-1}. We first claim that v must bring at least one unique subject z_1 (different from x), otherwise, u_k could unfollow v and follow y instead. Similarly, w must bring at least one unique subject z_2 (different from x and z_1). Then if there is no path from u_k to u_1, u_1 would unfollow x and follow u_k instead, so that he only loses one subject x but gains at least two subjects z_1 and z_2. □

The second Lemma aims at using the fact that users will tend to avoid redundant links at equilibrium.

Lemma 2 *Consider a strongly connected graph G with n nodes and m arcs (multiple arcs are allowed). If $m \geq 2n - 1$, then G contains a transitivity arc (i.e. an arc (s, t) such that there exists a directed path from s to t).*

Proof. We prove the result by induction on n. The hypothesis is true for $n = 1$ (a self-loop on vertex s is a transitivity arc for the empty directed path from s to s). We denote by $n(G)$ the number of nodes in the graph G and by $m(G)$ the number of edges in the graph G. Now consider $n > 1$ and assume that the property is true for any graph G' with $n(G') < n$. Consider a strongly connected graph G with n nodes containing no transitivity arc. Since $n \geq 2$, G must contain a circuit, i.e. an oriented cycle, with $k \geq 2$ nodes. The only arcs connecting two nodes of the circuit are the circuit arcs (otherwise, we would encounter a transitivity arc). Consider the graph G' obtained by contracting the circuit to one node. We have $m(G') = m(G) - k$ and $n(G') = n(G) - k + 1 < n$. Note that G' does not contain a transitivity arc either. Our induction hypothesis thus implies that $m(G') < 2n(G') - 1$. That is $m(G) - k < 2(n - k + 1) - 1$ or equivalently $m(G) < 2n - k + 1 \leq 2n - 1$ as $k \geq 2$.The property is thus satisfied for n. □

We are now ready to prove Proposition 1.

Proof.[of Proposition 1] Consider any equilibrium. Assume that a user u receives less than p subjects. u must be connected to some producer x by a path $x, u_1, \dots, u_k = u$. Consider the graph G' induced by users reachable from u_1 that receive less than p subjects. By Lemma 1, G' is strongly connected and all its users receive the same number $p' < p$ of subjects.

We claim that two users u and v of G' cannot follow the same producer y. As there exists a path from u to v, the link (y, v) would be redundant and v would be better off following some unreceived subject instead. Moreover, the fact that users in G' do not receive all subjects implies that they have spent all their budget of attention. We thus conclude that the number of edges in G', $m(G') = \sum_{u \in V(G')} \Delta_u - p'$. As the network is stable, there is no transitivity arc in G'. Lemma 2 thus implies $m(G') \leq 2n(G') - 2 \leq 2n(G')$, where $n(G')$ is the number of nodes in G'. We thus get $p' \geq \sum_{u \in V(G')} \Delta_u - 2n(G') = \sum_{u \in V(G')} (\Delta_u - 2)$.

First consider the case $p' \leq p - 2$. Suppose there exists a user $w \notin V(G')$. he cannot receive two subjects not received in G' otherwise u_1 would unfollow x and connect to w. As $\Delta_w \geq 3$, w can gather the p' subjects received in G' plus two others by connecting to one node in G' plus the two corresponding producers, a contradiction as this would increase his utility. We thus conclude that G' indeed contains all users, implying $p' \geq n(\overline{\Delta} - 2)$. Using (1), the utility of each user is at least $p' \geq \frac{\overline{\Delta}-2}{\overline{\Delta}-1} U^*$.

Finally, in all remaining cases to consider, all users receive at least $p - 1$ subjects. The utility of each user is thus at least $\frac{p-1}{p} U^* \geq \frac{\overline{\Delta}-2}{\overline{\Delta}-1} U^*$ as $p \geq \overline{\Delta} - 1$. □

3.2 Convergence of Dynamics

We have thus shown that stable configurations of self-organizing networks with homogeneous user interests are efficient. However, do network dynamics converge to an equilibrium ? The following proposition answers this question in the affirmative.

Proposition 2 *Any homogeneous flow game has an ordinal potential function, implying that selfish dynamics always converge to an equilibrium in finite time.*

Proof. Let n_i denote the number of users that receive i subjects and consider the sequence $(n_0, n_1, \ldots, n_p)$. We show that this sequence always decreases according to lexicographic ordering when users make selfish moves. The function $-\sum_{0 \leq i \leq p} n_i \, n^{p-i}$ is thus a potential function that will always increase until a local maximum is reached, proving convergence to an equilibrium.

Consider a user u that is receiving i subjects and that will make a selfish move to receive $j > i$ subjects instead. Note that there is no path from u to any other user receiving $k < i$ subjects. Therefore any change by u will not affect these users. Now consider any user v with $k \geq i$ subjects. If there is no path from u to v then u's selfish move does not affect v. If there is such a path, then v will now receive at least $j > i$ subjects. We thus now have $n_i - 1$ users receiving i subjects, and the sequence $(n_0, n_1, \ldots, n_p)$ has decreased according to lexicographic ordering. □

Our proof yields a very loose bound of n^{p+1} on convergence time. We leave as an open question whether exponential time of convergence can really arise. However,in the following proposition we show that a homogeneous flow game with at least 4 subjects, a user with budget of attention at least 2 and a user with budget of attention at least 3, is not equivalent to a congestion game. This rules out the possibility of using techniques similar to [8] to find equilibria in polynomial time, and more generally to easily bound convergence time.

Proposition 3 *Any homogeneous flow game with at least 4 subjects, a user with budget of attention at least 2 and a user with budget of attention at least 3, does not admit an exact potential function.*

Note that a game is equivalent to a congestion game iff it admits an exact potential function [17].

Proof. To show this, it is sufficient to exhibit a 4-cycle in the strategy space such that the sum of utility variations over the 4 moves is non-zero. (The variation of an exact potential potential function along the cycle would obviously be zero and would also have to be equal to that sum, leading to a contradiction as shown more formally in [17].) Without loss of generality, the game contains four producers $\{a, b, c, d\}$ and two users u, v with $\Delta_u \geq 2$ and $\Delta_v \geq 3$ as depicted in Figure 2. User u can adopt in particular strategy $A = \{(a, u)\}$ or $B = \{(b, u), (c, u)\}$. User v can adopt in particular strategy $C = \{(u, v), (b, v), (c, v)\}$ or $D = \{(u, v), (d, v)\}$. Consider the cycle $(A, C) \to (B, C) \to (B, D) \to (A, D) \to (A, C)$ where user u moves from strategy A to B increasing its utility by 1, then v moves from C to D and increases its utility by 1, then u moves back to A with a utility variation of -1, and finally v moves back to C increasing its utility by 1 again. The overall sum is thus $2 \neq 0$. $\square$

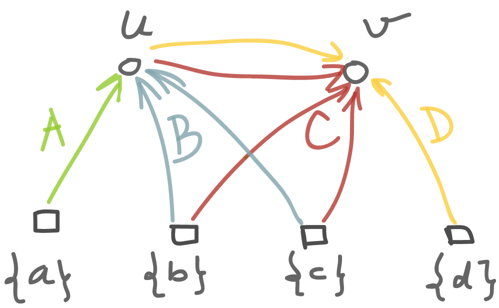


Figure 2: A 4-cycle $(A, C) \to (B, C) \to (B, D) \to (A, D) \to (A, C)$ in the strategy space.

Combining Proposition 1 and Proposition 2, we obtain:

Theorem 1 *In a homogeneous flow game where each user has budget of attention at least 3, less than p, and $\overline{\Delta}$ in average, selfish dynamics converge to an equilibrium* *such that the utility of a user is at least* $\frac{\overline{\Delta}-2}{\overline{\Delta}-1}U^*$ *where* U^* *is the optimal utility he can get, implying a price of anarchy of* $1 + 1/(\overline{\Delta} - 2)$ *at most.*

4 Heterogeneous interests

We now consider the more realistic case where users have differing sets of interests. To make the model even more general, we assume here that users weight independently topics. Let $W_u(s)$ denote the weight (or *value*) of topic s to user u. The objective of a user is now to maximize the sum of the values of subjects he receives. We will consider user-interest sets $S_u \subseteq P$ that include topics of sufficiently high value, that is $S_u = \{s : W_u(s) > 0\}$. The threshold ω_u serves as a filtering threshold, allowing more or less stringent filtering. Such user-specific weights for topics represent a natural expertise or focussed interest users may have on a subset of topics. (Note that the model presented previously corresponds to $W_u(s) = 1$ for $s \in S_u$, $W_u(s) = 0$ for $s \notin S_u$.) Here we assume user u is interested in a subset $S_u \subseteq P$ of topics. For the sake of simplicity, we assume $P = \cup_{u \in V(G)} S_u$. As a user may connect to other users whose interest sets differ from his own, he potentially receives subjects out of his interest set. The user may not have the resources to process and store this irrelevant information. We thus assume a natural filtering rule, where a user only retransmits subjects that are in his own interest set.

4.1 Price of Anarchy

We now show that the price of anarchy of such a system may be unbounded.

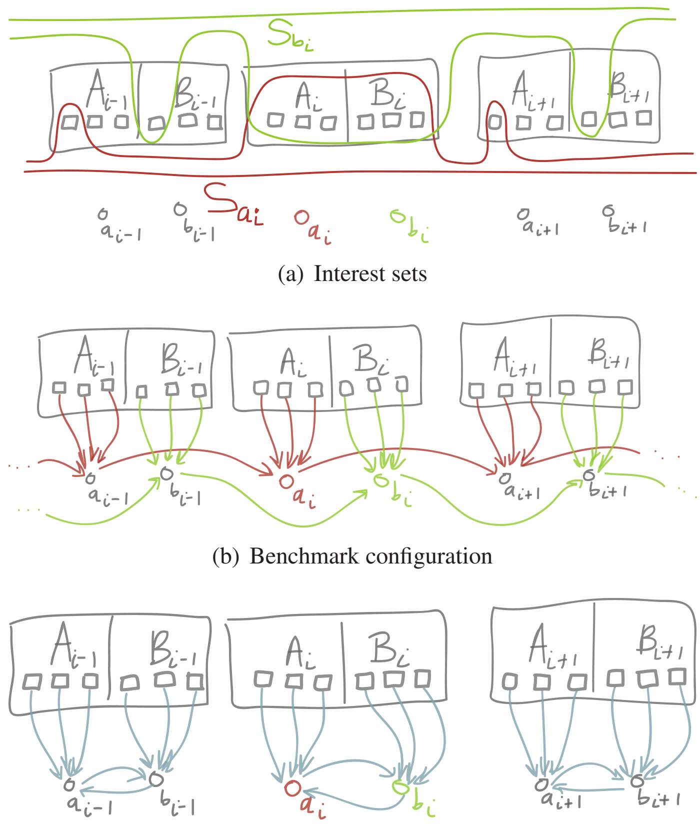
Proposition 4 *In a heterogeneous flow game, the price of anarchy can be arbitrarily large: specific choices yield a PoA of* $\Omega\left(\frac{n}{\Delta}\right)$.

Proof. We show the result through an example, illustrated in Figure 3. For integer k, consider a system with $n = 2k$ users having budget of attention $\Delta \geq 2$ each, and $p = 2(\Delta - 1)k$ producers. We distinguish two set of users $\{a_1, \ldots, a_k\}$ and $\{b_1, \ldots, b_k\}$. Similarly, the producers are partitionned into groups $\{A_1, \ldots, A_k\}$ and $\{B_1, \ldots, B_k\}$ where each A_i (resp. B_i) contains $\Delta - 1$ producers.

As illustrated in Figure 3(a), each user a_i has a value of 1 for each topic in $A_i \cup B_i$ and additionally the first element of each A_j for $j \neq i$. Similarly, each user b_i has a value of 1 for each topic in $A_i \cup B_i$ and additionally the first element of each B_j for $j \neq i$. Users have a value of zero for all other topics.

A benchmark configuration is shown in Figure 3(b), with two oriented rings, one for users a_i, $i = 1, \ldots, k$ and one for users b_i, $i = 1, \ldots, k$. User a_i is connected to a_{i-1} (with a_0 corresponding to a_k) and to all producers in A_i. User b_i is connected to b_{i-1} (with b_0 corresponding to b_k) and to all producers in B_i. The corresponding utility is $n(n/2 + \Delta - 2)$, so that the optimal global welfare $\mathcal{U}^*$ satisfies $\mathcal{U}^* \geq n^2/2$.

The configuration shown in Figure 3(c) is an equilibrium, where each user a_i (resp. b_i) connects to producers in A_i (resp. B_i) and to b_i (resp. a_i). The global utility here is $\mathcal{U} = n(2\Delta - 2) \leq 2n\Delta$, and the price of anarchy is thus at least $\frac{n}{4\Delta}$. $\square$



(a) Interest sets

(b) Benchmark configuration

(c) A Nash equilibrium configuration

Figure 3: Heterogeneous interest sets.

4.2 Convergence of dynamics

We have shown that the price of anarchy can be unbounded with respect to the number of users in some cases.

We now show that selfish dynamics do not even guarantee convergence to a Nash Equilibrium.

Proposition 5 *Selfish dynamics of a flow game with heterogeneous utilities may not converge.*

Proof. Consider the following scenario with six retransmitting users p_i, q_i, r_i, $i = 1, 2$, and two users u_1, u_2 each with degree $\Delta_i = 3$. The retransmitting users publish sets of topics as follows: $p_1 : \{a, b\}$, $p_2 : \{c, d\}$, $q_1 : \{x, y\}$, $r_1 : \{k, l\}$, $q_2 : \{x, k\}$, $r_2 : \{y, l\}$. The user-specific values are given in Table 1, where $\epsilon \ll 1$. As depicted in Figure 4, each agent u_i uses one connection to follow user p_i through whom he receives a total value of 4. He also connect to the other user u_{3-i} to receive another topic of value 2 from p_{3-i}. Now each user u_i must select between q_1, q_2, r_1 and r_2 for his third connection. We start with users u_1 and u_2 choosing q_1 and q_2 respectively. They thus receive $8 + \epsilon$ and $7 + 2\epsilon$ in total respectively. User u_2 then selects r_2, receiving $8 + \epsilon$, and this changes user u_1's utility to $7 + 2\epsilon$. However user u_1 can increase his utility by $1-\epsilon$, and does so by switching to r_1. This decreases u_2's utility by $1 - \epsilon$ since he doesn't receive x anymore (but now receives k). He can improve his utility by selecting q_1. Denote the state of the system by $(\mathcal{S}(u_1), \mathcal{S}(u_2))$ where $\mathcal{S}(u_i)$ is user u_i's strategy in selecting between q_i and r_i. Under selfish moves, the system may cycle as follows: $(q_1, q_2) \rightarrow (q_1, r_2) \rightarrow (r_1, r_2) \rightarrow (r_1, q_2) \rightarrow (q_1, q_2) \rightarrow (q_1, r_2) \rightarrow (r_1, r_2) \rightarrow \cdots$. □

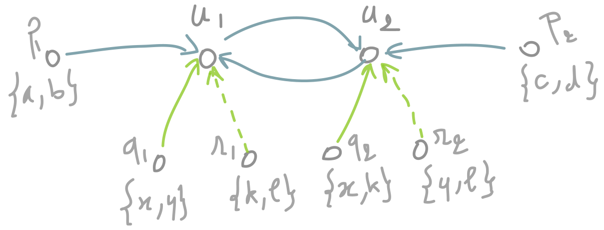


Figure 4: Instability with heterogeneous interest sets.

User\Topic	a	b	c	d	x	y	k	l
u_1	2	2	2	0	ϵ	1	1	ϵ
u_2	2	0	2	2	1	ϵ	ϵ	1

Table 1: User-specific values for topics.

With an arbitrary structure of user interest sets, we have thus shown that the price of anarchy may be unbounded, and dynamics may not converge. The question of determining if pure Nash equilibria exist is left open.

5 Structured interest sets

We now revisit the efficiency of social filtering in an heterogeneous scenario, where interest sets are no longer arbitrary but instead are organized according to a well behaved geometry. Specifically we assume the following model. A metric d is given on a set $P' \supseteq P$ of subjects. The interest set S_u of each user u then coincides with a *ball* $B(s_u, R_u)$ in this metric, specified by a *central subject* s_u and a *radius of interest* R_u. In other words, we assume that the value of a subject to a user is non-increasing in its distance from s_u. Specifically, we assume $W_u(s) = f(d(s_u, s))$ for $d(s_u, s) \leq R_u$, where $f(\cdot)$ is a non-increasing positive function, and $W_u(s) = 0$ otherwise. Without loss of generality, we can assume $P' = \{s_u : u \in V\} \cup P$ and $S_u = B(s_u, R_u) \cap P$. We shall first give conditions on the metric d and the sets S_u under which an efficient configuration exists. We will then introduce modified dynamics and filtering rules which guarantee stability, i.e. convergence to an equilibrium. A flow game where interest sets can be defined in this way is called a *metric flow* game.

The model can easily be generalized to more eclectic user interests where topics a user is interested in correspond to the disjoint union of a constant number of balls. We leave out the details of such generalizations so as to keep the focus of the paper. However, we include a brief discussion later in the section, in the context of Proposition 6.

5.1 Sufficient conditions for optimal utility

Consider the following properties of the interest set geometry.

1. γ-doubling: d is γ-doubling, i.e. for any subject s and radius R, the ball $B(s, R)$ can be covered by γ balls of radius $R/2$: there exists $I \subset S$ such that $|I| \leq \gamma$ and $B(s, R) \subset \cup_{t \in I} B(t, R/2)$.

2. r-covering: r is a covering radius, i.e. any subject $s \in P$ is at distance at most r from the central subject s_u of some user u with interest radius $R_u \geq r$.

3. (r, δ)-sparsity: there are at most δ subjects within distance r: $|B(s, r)| \leq \delta, \forall s$.

4. r-interest-radius regularity: for any users u, v with $d(s_u, s_v) < 3R_u/2 + r$, we have $R_v \geq R_u/2 + r$ (users with similar interests have comparable interest radii).

Property (1) is a classical generalization of dimension from Euclidean geometry to abstract metric spaces (an Euclidean space with dimension k is $2^{\Theta(k)}$-doubling). This is a natural assumption if user interests can be modeled by proximity in a hidden low-dimensional space. Property (2) states that all subjects are within distance r from some user's center of interest and can thus be seen as an assumption of minimum density of users' interests over the whole set P of available subjects. Property (3) puts an upper bound on the density of subjects. In other words, we assume a level of granularity under which we do not distinguish subjects. Property (4) is another form of regularity assumption, requiring some smoothness in the radii of interests of nearby users. This may be the most debatable assumption, for instance if we consider the case of an expert next to an amateur. However, if we assume that a topic is split into several subjects according to the level of expertise required to understand the corresponding news, the assumption becomes more natural as an expert is still interested in related subjects (with lower level of understanding) and an amateur still has some focus if the correct number of levels is considered.

We now show that an optimal solution exists, i.e. one in which each user receives all subjects in his interest set, as soon as his budget of attention is at least $\gamma\delta + \gamma^2 \log \frac{R_m}{r}$ where R_m is the maximum radius of interest over all users. This will be a direct consequence of the following proposition.

Proposition 6 *Consider a metric flow game satisfying the γ-doubling, r-covering, (r, δ)-sparsity and r-interest-radius regularity assumptions. If in addition each user u has a budget of attention at least $\gamma\delta + \gamma^2 \log \frac{R_u}{r}$, then there exists a collection of user strategies allowing each user u to receive all subjects in S_u.*

This result can easily be extended to the case where each user interest set is given by a disjoint union of balls (the number of balls being at most a constant b). It suffices to repeat the construction of the proof for each ball, resulting in a factor b in the resulting required budget of attention. The assumptions have to be slightly modified so that any subject is covered by some ball of a user (in the covering assumption) and that two nearby balls have comparable radii (in the regularity assumption).

Proof. We define the ball $B_{u,i} := B(s_u, \min(R_u, 2^i r))$ for each user u and each integer $i \geq 0$. The construction to follow will ensure that u collects all subjects in $B_{u,i}$ through a set $N_{u,i}$ of contacts such that $B_{u,i} \subset \cup_{v \in N_{u,i}} B_{v,i-1}$.

We first define $N_{u,1} = \{p_s : s \in B_{u,1}\}$. Now, for $2 \leq i \leq \lceil \log \frac{R_u}{r} \rceil$, the γ-doubling assumption implies that $B_{u,i}$ can be covered by at most γ^2 balls of radius $2^{i-2}r$: there exists a set $L_{u,i}$ of at most γ^2 subjects such that $B_{u,i} \subset \cup_{s \in L_{u,i}} B(s, 2^{i-2}r)$. From the r-covering assumption, we can then define a set $N_{u,i}$ of at most γ^2 users such that each $s \in L_{u,i}$ is at distance at most r from some s_v with $v \in N_{u,i}$. We then have $B_{u,i} \subset \cup_{v \in N_{u,i}} B(s_v, 2^{i-2}r + r)$. Without loss of generality, we can assume that for each $s \in L_{u,i}$, $B(s, 2^{i-2}r)$ intersects $B_{u,i}$ (otherwise s can safely be removed from $L_{u,i}$ as it does not cover anything useful). We thus have $d(s_u, s) \leq R_u + 2^{i-2}r < 3R_u/2$ (note that $2^{i-1}r < R_u$ as $i \leq \lceil \log \frac{R_u}{r} \rceil$). For $v \in N_{u,i}$ such that $d(s, s_v) \leq r$, we then have $d(s_u, s_v) < 3R_u/2 + r$. From the r-interest-radius regularity, we then deduce $R_v \geq R_u/2 + r > 2^{i-2}r + r$, implying

$\min(R_v, 2^{i-1}r) \geq 2^{i-2}r + r$. The ball $B_{v,i-1}$ thus contains $B(s_v, 2^{i-2}r + r) \supset B(s, 2^{i-2}r)$. Together with the definition of $L_{u,i}$, this proves $B_{u,i} \subset \cup_{v \in N_{u,i}} B_{v,i-1}$.

The connection graph G results from connecting each user u to all contacts in the set $\cup_{1 \leq i \leq \lceil \log \frac{R_u}{r} \rceil} N_{u,i}$.

Flow correctness: We show by induction on i that each user u receives all subjects in $B_{u,i}$. The direct connection to producers for subjects in $B_{u,1}$ ensures this for $i = 1$. For $i > 1$, the induction hypothesis implies that each user $v \in N_{u,i}$ receives all subjects in $B_{v,i-1}$. From $B_{u,i} \subset \cup_{v \in N_{u,i}} B_{v,i-1}$, we conclude that u will receive news about subjects in $B_{u,i}$ from its contacts in $N_{u,i}$. As $S_u = B_{u, \lceil \log \frac{R_u}{r} \rceil}$, we finally know that u receives all subjects in S_u.

In-degree bound: First, we have $|N_{u,1}| \leq \gamma\delta$. This comes from the fact that $B_{u,1}$ is included in at most γ balls of radius r from the γ-doubling assumption, and each of these balls contains at most δ subjects from the (r, δ)-sparsity assumption. Second, we have already seen that $|N_{u,i}| \leq \gamma^2$ for $2 \leq i \leq \lceil \log \frac{R_u}{r} \rceil$. We thus obtain the bound $\gamma\delta + \gamma^2 \left(\lceil \log \frac{R_u}{r} \rceil - 1 \right) < \gamma\delta + \gamma^2 \log \frac{R_u}{r}$. □

The core of the construction consists in covering a a given ball radius of $2^i r$ with a set of γ balls of radius $2^{i-1}r$. As a covering set of γ^2 balls can be computed through a simple greedy covering algorithm [11], a solution where the required budget of attention is within a factor γ from the bound of Proposition 6 can thus be computed in polynomial time.

As previously mentioned, a budget of attention of $\Delta = \gamma\delta + \gamma^2 \log \frac{R_m}{r}$ per user is thus enough for maximum utility. This scales logarithmically in R_m, while under the assumptions of the theorem one can arrange interest sets to have size polynomial in R_m (take for example interests to be regularly placed on a lattice). Thus this configuration gives substantial savings in comparison to one where users would connect directly to all their subjects.

Clearly the configuration graph identified in this theorem is an equilibrium: as maximum utility is reached, no user can increase its utility by reconnecting. We now study conditions that guarantee convergence of dynamics.

5.2 Sufficient conditions for stability

We first define two rules regarding republication of subjects received and reconnections.

1. Expertise-filtering rule: when a user u is connected to a user v, u only receives subjects s such that $d(s_v, s) \leq d(s_u, s)$.

2. Nearest-subject rule for re-connection: when reconnecting, each user u gives priority to subjects that are closer to s_u: a new subject s is gained by u so that no subject t with $d(s_u, t) < d(s_u, s)$ is lost. (On the other hand, any subject t with $d(s_u, t) > d(s_u, s)$ can be lost.)

Rule 1 can be interpreted as follows. The center of expertise of a user is the same as its center of interest, and the distance d also captures expertise of users about subjects, in that u is more expert than v on subject s if and only if $d(s_u, s) \leq d(s_v, s)$. The rule then amounts to a sanity check where u discards news from sources that have less expertise than himself on the subject. We capture with the following slight variation of the model. A flow game *with expertise-filtering* is a flow game where reception of a subject s by user u occurs only when there exists a directed path $s = u_0, \ldots, u_k = u$ from s to u such that for each $1 \leq i < k$, $s \in S_{u_i}$ (i.e. $d(s_{u_i}, s) \leq R_{u_i}$) and $d(s_{u_i}, s) \leq d(s_{u_{i+1}}, s)$.

Rule 2 states that a user u prefers to receive a subject he is more interested in (i.e. closer to s_u) rather than any number of subjects that are less interesting. A flow game is denoted to be *with nearest-subject priority* if the utility function of each user u is defined by $U_u(F) = \max\{R : u \text{ receives all } s \in B(s_u, R)\}$.

Proposition 7 *Any metric flow game with expertise-filtering and nearest-subject priority has an ordinal potential function, implying that selfish dynamics always converge to an equilibrium in finite time.*

The proof shows the existence of an ordinal potential function. As in the previous section, the bound on convergence time implied by the above proof is very loose. We leave open the question of determining better bounds or faster convergence conditions.

Proof. Consider the set $\mathcal{D} = \left\{d(s,t) : s, t \in P'^2\right\}$ of all possible distances. Let $r_1, \ldots, r_m$ denote all elements of $\mathcal{D}$ sorted in increasing order (i.e. $r_1 < \cdots < r_m$). Let n_i denote the number of pairs (u, s) such that $d(s_u, s) = r_i$ and u receives s. Consider the tuple $(n_1, \ldots, n_m)$. When a user u makes a selfish move, it increases its utility by receiving a new subject s. Let i denote the index such that $d(s_u, s) = r_i$. Any lost subject t must satisfy $d(s_u, t) > d(s_u, s)$ by the nearest-subject rule. If a lost subject t was received by some user v through a path from u to v, we have $d(s_v, t) \geq d(s_u, t)$ by the expertise-filtering rule. We thus deduce $d(s_v, t) > d(s_u, s)$, implying that n_j can decrease only for $j > i$. The tuple $(n_1, \ldots, n_m)$ thus increases according to the lexicographical order after any selfish move. The function $\sum_{0 \leq i \leq m} n_i\,(n+p)^{2(m-i)}$ is thus a potential function that will always increase until a local maximum is reached, proving convergence to an equilibrium. □

We are now ready to prove the following:

Theorem 2 *Consider a metric flow game with expertise-filtering and nearest-subject priority that satisfies the γ-doubling, r-covering, (r,δ)-sparsity and r-interest-radius regularity assumptions. If in addition each user u has budget of attention at least $\gamma\delta + \gamma^2 \log \frac{R_u}{r}$, selfish dynamics converge to an equilibrium where each user u receives all subjects in S_u, implying that the price of anarchy is then 1.*

Proof. Consider a configuration where some users do not receive some subject in their interest ball. Let (u,s) be a user-subject unsatisfied pair such that $d(s_u,s)$ is minimal. Consider the smallest integer i such that $d(s_u,s) \leq 2^i r$ holds. According to the construction of Proposition 6, user u can receive all subjects in $B_{u,i} = B(s_u, \min(R_u, 2^i r))$ as long as every user v receives all subjects in his ball of radius $\min(R_v, 2^{i-1}r)$ which is the case according to the choice of the pair (u,s). Note that this construction follows the expertise filtering rule as each subject at distance greater than $2^{i-1}r$ is retrieved through a user at distance at most $2^{i-1}r$ from the subject. User u can retrieve $B_{u,i}$ using at most $\gamma\delta+\gamma^2(i-1)$ connections. The configuration is thus unstable as long as $\Delta_u \geq \gamma\delta + \gamma^2(i-1)$ which is the case for $\Delta_u \geq \gamma\delta + \gamma^2 \log \frac{R_u}{r}$. Since the system must stabilize to some equilibrium according to Proposition 7, every user u must receive all news about subjects in S_u in that stable configuration. □

Interestingly, the above proof implies that the convergence is fast: as soon as all users receive their ball of radius $2^{i-1}r$, one reconnection by each user will allow him to receive his ball of radius $2^i r$ (expertise-filtering and nearest-subject priority ensure that other users will not lose subjects at distance less than $2^i r$). Convergence is thus achieved after $\log \frac{R_m}{r}$ rounds where each round consists in letting each user reconnect once (or more).

6 Concluding remarks

We have shown that a flow game can have complex dynamics that may not converge. However, we can prove convergence to efficient equilibrium for both homogeneous flow games (with very weak assumptions) and metric flow games (with more technical assumptions). While our proofs give exponential bounds on convergence time in general, we get linear convergence time (up to a logarithmic factor) for structured interest set with expertise-filtering and nearest-subject priority, showing that understanding the structure of interests and its relation to forwarding mechanisms is a key aspect of information flow in social networks. Direct follow up of this work concerns the study of the speed of convergence in general and the characterization of flow games having pure Nash equilibria.

Our model makes several simplifying assumptions. A natural generalization would be to consider a real-valued cost of attention for establishing a link (v,u)

instead of a unitary cost. The cost of establishing link (v, u) could typically be a function of S_u and S_v. A natural cost taking into account the attention required to filter out uninteresting content would then be $c(v, u) = \frac{|S_v|}{|S_u \cap S_v|}$, for example.

A dual variant of our model could be to consider that every user gathers all the subjects he is interested in while he tries to minimize the required cost of attention. We could also mix both models, using utility functions combining coverage of interest set and cost of attention (the function being increasing in the number of interesting subjects received and decreasing in the costs of attention of the formed links).

In that context, we believe the two following directions are promising for efficient social dissemination. First, incentive mechanisms, e.g. reputation counters maintained by users, or payments between users, may considerably augment the performance of self-organizing social flows. Second, more elaborate content filtering between contact-follower pairs may also lead to substantial improvements. We have already introduced expertise filtering, which could translate into implementable mechanisms in existing social networking platforms. More generally there appears to be a rich design space of filtering rules based on combinations of interests and expertise.

References

[1] I. Abraham, D. Malkhi, and O. Dobzinski. LAND: stretch $(1+\epsilon)$ locality-aware networks for DHTs. In J. I. Munro, editor, *SODA*, pages 550–559. SIAM, 2004.

[2] E. Anshelevich, A. Dasgupta, J. Kleinberg, E. Tardos, T. Wexler, and T. Roughgarden. The price of stability for network design with fair cost allocation. In *Proceedings of the 45th Annual IEEE Symposium on Foundations of Computer Science*, FOCS '04, pages 295–304, Washington, DC, USA, 2004. IEEE Computer Society.

[3] V. Bala and S. Goyal. A noncooperative model of network formation. *Econometrica*, 68(5):1181–1229, 2000.

[4] X. Bei, W. Chen, S.-H. Teng, J. Zhang, and J. Zhu. Bounded budget betweenness centrality game for strategic network formations. *Theor. Comput. Sci.*, 412(52):7147–7168, Dec. 2011.

[5] M. Cha, H. Haddadi, F. Benevenuto, and P. K. Gummadi. Measuring user influence in twitter: The million follower fallacy. In W. W. Cohen and S. Gosling, editors, *ICWSM*. The AAAI Press, 2010.

[6] K. Clarkson. Nearest neighbor queries in metric spaces. *Discrete & Computational Geometry*, 22(1):63–93, 1999.

[7] A. Fabrikant, A. Luthra, E. Maneva, C. H. Papadimitriou, and S. Shenker. On a network creation game. In *Proc. ACM PODC*, pages 347–351, 2003.

[8] A. Fabrikant, C. H. Papadimitriou, and K. Talwar. The complexity of pure nash equilibria. In L. Babai, editor, *STOC*, pages 604–612. ACM, 2004.

[9] P. Fraigniaud, E. Lebhar, and L. Viennot. The inframetric model for the internet. In *Proceedings of the 27th IEEE International Conference on Computer Communications (INFOCOM)*, pages 1085–1093, Phoenix, 2008.

[10] A.-T. Gai, D. Lebedev, F. Mathieu, F. De Montgolfier, J. Reynier, and L. Viennot. Acyclic Preference Systems in P2P Networks. In *Proc. Euro-Par*, 2007.

[11] S. Har-Peled and M. Mendel. Fast construction of nets in low dimensional metrics, and their applications. In J. S. B. Mitchell and G. Rote, editors, *Symposium on Computational Geometry*, pages 150–158. ACM, 2005.

[12] M. Jackson. *Social and Economic Networks*. Princeton University Press. Princeton University Press, 2010.

[13] B. Jiang, N. Hegde, L. Massoulié, and D. Towsley. How to optimally allocate your budget of attention in social networks. In *Proc. IEEE Infocom*, 2013.

[14] D. R. Karger and M. Ruhl. Finding nearest neighbors in growth-restricted metrics. In J. H. Reif, editor, *STOC*, pages 741–750. ACM, 2002.

[15] N. Laoutaris, L. J. Poplawski, R. Rajaraman, R. Sundaram, and S.-H. Teng. Bounded budget connection (BBC) games or how to make friends and influence people, on a budget. In *Proceedings of the twenty-seventh ACM symposium on Principles of distributed computing*, PODC '08, pages 165–174, New York, NY, USA, 2008. ACM.

[16] L. Massoulié and A. Twigg. Rate-optimal schemes for peer-to-peer live streaming. *Journal of Performance Analysis*, 2008.

[17] D. Monderer and L. Shapley. Potential games. *Games and Economic Behavior*, pages 124–143, 1996.

[18] N. Nisan, T. Roughgarden, É. Tardos, and V. V. Vazirani, editors. *Algorithmic Game Theory*. Cambridge Univ Press, 2007.

[19] C. G. Plaxton, R. Rajaraman, and A. W. Richa. Accessing nearby copies of replicated objects in a distributed environment. *Theory Comput. Syst.*, 32(3):241–280, 1999.

[20] R. W. Rosenthal. A class of games possessing pure-strategy Nash equilibria. *International Journal of Game Theory*, 2:65–67, 1973.

[21] T. Roughgarden. Potential functions and the inefficiency of equilibria. *Proceedings of the International Congress of Mathematicians (ICM)*, 3:1071–1094, 2006.

[22] H. A. Simon. Designing organizations for an information rich world. In M. Greenberger, editor, *Computers, communications, and the public interest*, pages 37–72. The Johns Hopkins Press, Baltimore, 1971.