



HAL
open science

Self-Organizing Flows in Social Networks

Nidhi Hegde, Laurent Massoulié, Laurent Viennot

► **To cite this version:**

Nidhi Hegde, Laurent Massoulié, Laurent Viennot. Self-Organizing Flows in Social Networks. [Research Report] 2012. hal-00761046v1

HAL Id: hal-00761046

<https://inria.hal.science/hal-00761046v1>

Submitted on 4 Dec 2012 (v1), last revised 27 Feb 2015 (v3)

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

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 the interested users. Our motivation in this paper is to measure to what extent self-organization of the 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 in contrast, for the more realistic case of heterogeneous interests, that convergence, if it occurs, may lead to information dissemination that can be arbitrarily inefficient, as captured by an unbounded “price of anarchy”.

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

1.1 Motivation

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.

*Part of this work was done while at Technicolor.

†Supported by the Inria project-team “Gang”.

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 her interest. On the other hand, a user pays with her attention: filtering interesting information from spam (i.e. information that does not fall in her topics of interest) incurs a cost. Users must therefore trade-off topic coverage against attention cost. As pointed out by Simon [20], 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 decision on creating connections. To model this, we introduce a *flow* game in network formation where some users produce news about specific topics and each user is interested in receiving all news about a set of topics specific to her. 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 her 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. This model could also be of interest in the context of peer-to-peer streaming and file sharing or publish/subscribe applications.

1.2 Our results

We restrict ourselves to the case where each incoming link incurs a unitary attention cost and where the budget of attention of each user is captured by an integral value. This assumption is justified for almost all the results of the paper where users connect to users sharing at least a constant fraction of topics of interest (the cost of reading news arriving from various connections can then be considered to be roughly the same). The utility of a user is measured by the number of topics of her interest that she receives. We additionally assume that a user produces news about one topic at most even if she redistributes other topics. This is coherent with an empirical study of twitter traces [4] where it is shown that ordinary users (as opposed to celebrities or newspapers) can gain influence by concentrating on a single topic. Although we make several simplifying assumptions, we believe our model grasps sufficient complexity and tackles the main phenomena.

We derive conditions where selfish dynamics (when each user independently and repeatedly tries to increase her own utility) converge to a pure Nash equilibrium, that is a state where no user can receive more topics by changing her connections while other users keep the same connections. (In the sequel we will simply speak of equilibrium for 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, that is the ratio of the global welfare (measured as the sum of user utilities) at an optimal solution compared to the global welfare at the worst equilibrium.

More precisely, we first consider the homogeneous case where all users are interested in the same set of topics. In this case, an optimal solution can easily be constructed by forming a ring between users with budget of attention at least 2. We show that the homogeneous game is not an exact potential game, that is, it does not admit an exact potential function as defined by [15]. (Equivalently it is not a congestion game [18, 15].) In particular, this rules out the possibility of bounding the price of anarchy based on classical techniques using potential functions as described in [19]. Nevertheless, the game is an ordinary potential game (we exhibit a potential function that decreases under a selfish move) implying that selfish dynamics converge to an equilibrium in finite time. Additionally, we prove that the price of anarchy is bounded by 2 as soon as budget of attention is at least 3 for each user. More precisely, the price of anarchy is bounded by $1 + \frac{1}{\bar{\Delta}-2}$ where $\bar{\Delta}$ is the average budget of attention. We indeed prove the stronger result that the utility of a user at equilibrium is within a constant factor (approaching 1 as $\bar{\Delta}$ increases) of the maximal utility she can get in any configuration.

We then consider the heterogeneous case where each user has her own set of topics of interests. The situation is then much more complex as computing an optimal solution may be NP-hard: one can easily reduce the max-cover problem to it. We leave open the question of determining if pure Nash equilibria always exist in this context. However, we exhibit particular configurations where highly inefficient topologies may arise, i.e. equilibria with linear price of anarchy. Thus in general there is scope for improving the performance of social networks by augmenting self-organization with suitable mechanisms.

However, we show that this is not necessary when users' interests exhibit a particular structure, captured by a metric space. We consider the case where the topics form a subset of a metric space and where a user is interested in the topics falling in a given ball of the metric space. The ball can be seen as a specific domain of interest. (More realistically, we can consider that a user has several domains of interest by adapting the model for using a union of balls instead of just one ball.) With a slight modification to our model which is natural in this case, we can again exhibit an ordinal potential function implying that selfish dynamics converge to an equilibrium in finite time. Assuming that the metric space has low doubling dimension and similar related properties, we further show convergence to an equilibrium where users obtain all the information they are interested in, provided their budget of attention is logarithmic in the number of topics they are interested in. This case yields a price of anarchy of 1. The low doubling dimension occurs in particular when each user interests and each topic can be captured by a vector in a hidden euclidean space with small dimension, and where a user is interested in the topics whose vectors are within a certain distance from the user vector.

1.3 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 is thus in a follow-up on the large amount of work in network formation games. However, to the best of our knowledge, the this objective of efficient information dissemination 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 [11] 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.

Network creation games in the context of the Internet have been considered [16], 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 [6], or ensuring connection to a given subset of nodes [2]. We consider a bound of edge costs, in the form of a limit on the number in-edges at each node, and further, we focus on connections that allow flows of information.

Interestingly, bounded budget network formation games have already been considered. Bounded budget connection games [13] 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 [3] 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.

The notion of connecting to users that can provide a content flow of interest is similar to peer-to-peer live streaming systems [14]. Unlike peer-to-peer streaming, our model has download constraints (in the form of budget of attention) and 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 [9]. As a generalization of the stable marriages problem, those systems consider configurations of undirected edges based on mutual acceptance of an edge where unilateral decision is more suitable in our model. Our model is more intricate in the sense that connections are based not only on preferences (and affinity with other users) 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 [17] that can grasp general situations 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 [5], network construction [1], closest server selection [12], etc. Doubling metrics have mainly been used to model distances in networks such as Internet [8].

1.4 Organization of the paper

Section 2 introduces the model. We study the case of homogeneous interests in Section 3. We first bound the price of anarchy in Subsection 3.1 and then consider convergence under selfish dynamics and potential functions in Subsection 3.2. The heterogeneous case in its full generality is considered in Section 4 which details some negative results. Section 5 is dedicated to the restrictive scenario where users' interests are captured by 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 produces 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 $S : V \rightarrow P$ is a function associating to each user u its interest set $S_u \subseteq P$, and $\Delta : V \rightarrow \mathbb{N}$ 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 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 . 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 . 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. (Dynamics converge when any sequence of selfish moves leads to an 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.

An ordinary (or general [7]) potential function [15] is a function $f : \mathcal{F} \rightarrow \mathbb{R}$ such that $\text{sign}(f(F') - f(F)) = \text{sign}(U_u(F') - U_u(F))$ 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 [15] who show that it is tightly related to the notion of a congestion game [18]. Potential functions are classically used to show convergence of dynamics and to bound price of anarchy [7, 19].

3 Homogeneous interests

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

convergence of dynamics.

3.1 Bounding the 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, it cannot obtain more subjects than the aggregate budget of attention of all users, that is $\sum_{u \in V(G)} \Delta_u$, which we also denote by $n\bar{\Delta}$. However, this bound is attained only when all users have budget of attention 1. When there are at least two users with budget at least 2 and less than p (we will restrict to that more interesting case), one can easily see that the optimal solution 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 connection 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 $\sum_{u \in V(G)} \Delta_u - 1$. In that case (budget at least 2 and less than p for at least two users), we have thus established the following upper bound on the maximal utility U^* a user can get:

$$U^* \leq \min(p, n(\bar{\Delta} - 1)). \quad (1)$$

We now consider a distributed setting where each user selfishly rewires her incoming connections if she can improve her utility, i.e., if this allows her 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 $3 \leq \Delta_u \leq p$ for every user $u \in V$ of an homogeneous flow game. Then for any equilibrium the utility of a user is at least $\frac{\bar{\Delta}-2}{\bar{\Delta}-1}U^*$ where U^* is the optimal utility she can get. The price of anarchy is thus at most $1 + 1/(\bar{\Delta} - 2)$, approaching 1 for large $\bar{\Delta}$.*

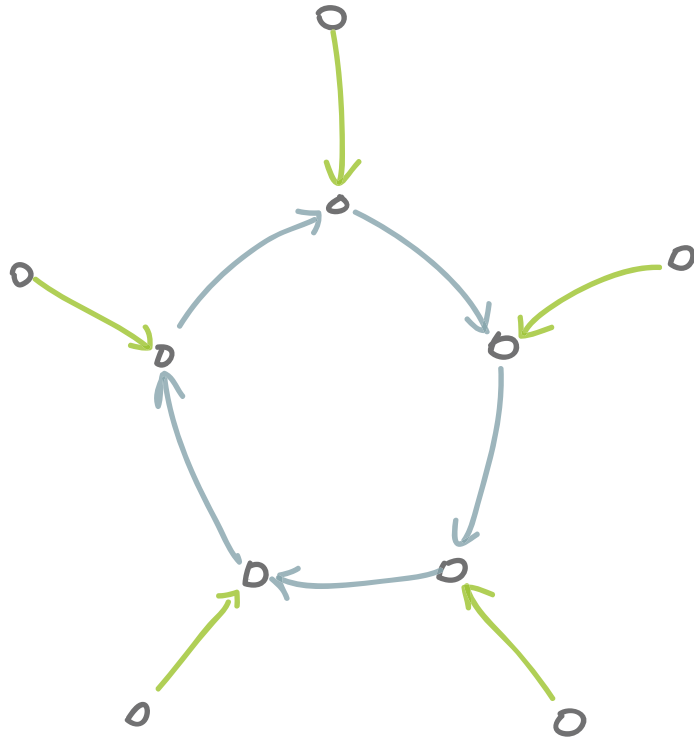
We first note that the above theorem is tight in the sense that high price of anarchy can arise for $\bar{\Delta} = 2$ as shown in Figure 1. In this particular case, a doubly linked chain forms a Nash equilibrium gathering only two subjects in total while an oriented cycle gathers n subjects. The price of anarchy is thus $n/2$.

We now establish two lemmas before proving Proposition 1. We first establish the existence of strongly connected components in any stable network.

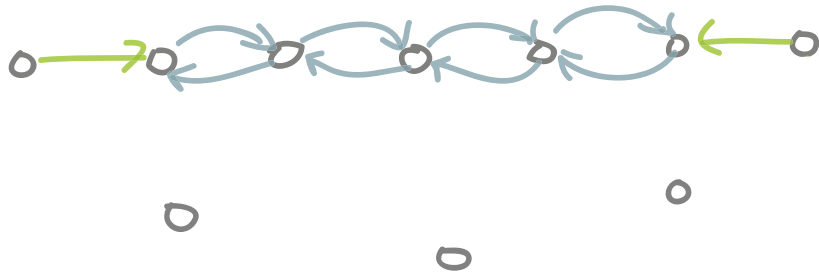
Lemma 1 *If an equilibrium is reached such that there exists a path x, u_1, \dots, u_k where x is a producer, u_k has in-degree bound $\Delta_{u_k} \geq 3$ and a producer y is not received by u_k , then there is a path from u_k to u_1 .*

Proof. 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 , otherwise, u_k could unfollow v and follow y instead. Similarly, w must bring at least one unique subject z_2 . Then if there is no path from u_k to u_1 , u_1 would unfollow x and follow u_k instead, so that she only loses one subject x but gains at least two subjects z_1 and z_2 . \square

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



(a) Benchmark configuration



(b) A Nash equilibrium configuration

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

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 $n = 1$. 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 . \square

We are now ready to prove Proposition 1.

Proof.[of Proposition 1] Consider any equilibrium. Assume that some user u receives less than p subjects. u must be connected to some producer x by a path $x, u_1, \dots, u_k = u$ (eventually $k = 1$). 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 follow 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 $m(G') = \sum_{u \in V(G')} \Delta_u - p'$. As the network is stable, there is no transitivity arc in G' . Lemma 2 thus implies $m(G') \leq 2n(G') - 2 \leq 2n(G')$. 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 $w \notin V(G')$. She 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 her utility. We thus conclude that G' indeed contains all users, implying $p' \geq n(\bar{\Delta} - 2)$. Using (1), the utility of each user is at least $p' \geq \frac{\bar{\Delta}-2}{\bar{\Delta}-1} U^*$.

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

3.2 Convergence to equilibrium and potential functions

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 the sequence (n_0, n_1, \dots, 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, \dots, n_p) has decreased according to lexicographic ordering. \square

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, we show that an 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 [7] to find equilibria in polynomial time, and more generally to easily bound convergence time.

To prove this, we show that the game does not admit an exact potential function (which is equivalent to not being equivalent to a congestion game [15]). 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 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 [15].) 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 $(A, C) \rightarrow (B, C) \rightarrow (B, D) \rightarrow (A, D) \rightarrow (A, C)$ 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 $2 \neq 0$.

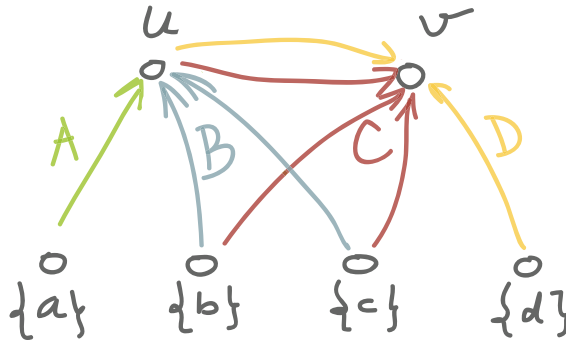


Figure 2: A 4-cycle $(A, C) \rightarrow (B, C) \rightarrow (B, D) \rightarrow (A, D) \rightarrow (A, C)$ in the strategy space.

Combining Proposition 1 and Proposition 2, we obtain:

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

4 Heterogeneous interests

We now consider the more realistic case where users have differing sets of interests. We assume user u is interested in a subset $S_u \subseteq P$ of subjects. We will show that the price of anarchy of such a system may be unbounded.

Proposition 3 *In an 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 $\{a_1, \dots, a_k\}$ and $\{b_1, \dots, b_k\}$. Similarly, the producers are partitioned into groups $\{A_1, \dots, A_k\}$ and $\{B_1, \dots, B_k\}$ where each A_i (resp. B_i) contains $\Delta - 1$ producers.

As illustrated in Figure 3(a), each user a_i is interested in $A_i \cup B_i$ and additionally the first element of each A_j for $j \neq i$. Similarly, each user b_i is interested in $A_i \cup B_i$ and additionally the first element of each B_j for $j \neq i$.

A benchmark configuration is shown in Figure 3(b), with two oriented rings, one for users a_i , $i = 1, \dots, k$ and one for users b_i , $i = 1, \dots, k$. 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 $\mathcal{U}^* \geq n^2/2$.

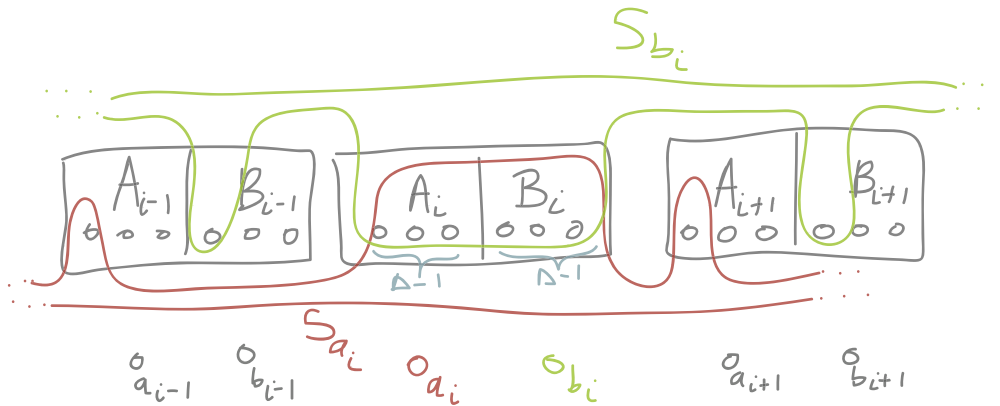
On the other hand, the configuration shown in Figure 3(c) is an equilibrium, where each user a_i (resp. b_i) connects to producers in A_i (resp. B_i) and to b_i (resp. a_i). The global utility here is $\mathcal{U} = n(2\Delta - 2) \leq 2n\Delta$, and the price of anarchy is thus at least $\frac{n}{4\Delta}$.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

We have shown that the price of anarchy can be unbounded with respect to the number of users in some cases. The question of determining if pure Nash equilibria exist is left open. We conjecture that selfish dynamics may not converge.

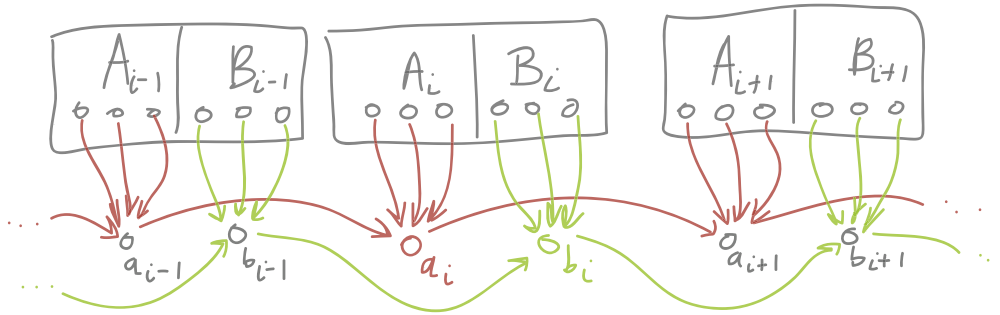
5 Structured interest sets

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 . 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*.

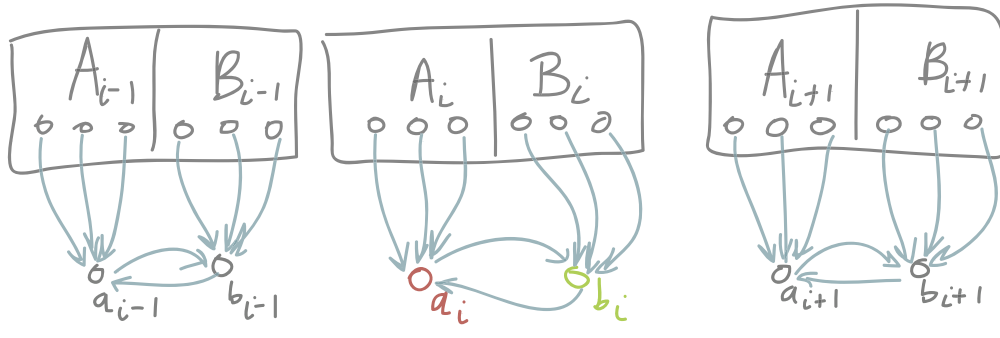
The model can easily be generalized to more eclectic user interests where topics a user is interested in correspond to the disjoint union of a constant number of balls. We leave out the details of such generalizations so as to keep the focus of the paper. However, we include a brief discussion later in the section, in the context of Proposition 4.



(a) Interest sets



(b) Benchmark configuration



(c) A Nash equilibrium configuration

Figure 3: Heterogeneous interest sets.

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 : $|B(s, r)| \leq \delta$ for all 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 her interest set, as soon as her budget of attention Δ_u satisfies $\Delta_u \geq \gamma\delta + \gamma^2 \log \frac{R_m}{r}$ where R_m is the maximum radius of interest over all users. This will be a direct consequence of the following proposition.

Proposition 4 *If a metric flow game satisfies 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 strategies such that each user u receives 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. For each user u and each integer $i \geq 0$, define the ball $B_{u,i} := B(s_u, \min(R_u, 2^i r))$. 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 $\gamma\delta + \gamma^2 (\lceil \log \frac{R_u}{r} \rceil - 1) < \gamma\delta + \gamma^2 \log \frac{R_u}{r}$. \square

A set of γ^2 balls of radius $2^{i-1}r$ sufficient to cover a given ball radius of $2^i r$ can be computed through a simple greedy covering algorithm [10]. A solution where the required budget of attention is within a factor γ from the bound of Proposition 4 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 selfish 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 $d(s_v, s) \leq d(s_u, u)$.
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 herself 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 $s = u_0, \dots, u_k = u$ 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 $d(s_{u_i}, s) \leq d(s_{u_{i+1}}, s)$ (expertise filtering).

Rule 2 simply states that a user u prefers to receive a subject it 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 $U_u(F) = \max \{R : u \text{ receives all } s \in B(s_u, R)\}$ (we simply choose a function naturally reflecting the rule).

Proposition 5 *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.*

Proof. Consider the set $\mathcal{D} = \{d(s, t) : s, t \in P^{l^2}\}$ of all possible distances. Let r_1, \dots, r_m denote all elements of \mathcal{D} sorted in increasing order (i.e. $r_1 < \dots < r_m$). 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, \dots, 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, \dots, 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. \square

Again 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.

We are now ready to prove the following:

Theorem 2 *A metric flow game with expertise-filtering and nearest-subject priority 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 4, 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 her 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 5, every user u must receive all news about subjects in S_u in that stable configuration. \square

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). Direct follow up of this work concerns the study of the speed of convergence and the characterization of flow games having pure Nash equilibria.

Our model makes several simplifying assumptions. We believe that several of them could be alleviated. 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 she is interested in while she 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] Ittai Abraham, Dahlia Malkhi, and Oren Dobzinski. Land: stretch $(1 + \epsilon)$ locality-aware networks for dhts. In J. Ian Munro, editor, *SODA*, pages 550–559. SIAM, 2004.
- [2] Elliot Anshelevich, Anirban Dasgupta, Jon Kleinberg, Eva Tardos, Tom Wexler, and Tim 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] Xiaohui Bei, Wei Chen, Shang-Hua Teng, Jialin Zhang, and Jiajie Zhu. Bounded budget betweenness centrality game for strategic network formations. *Theor. Comput. Sci.*, 412(52):7147–7168, December 2011.

- [4] Meeyoung Cha, Hamed Haddadi, Fabrício Benevenuto, and P. Krishna Gummadi. Measuring user influence in twitter: The million follower fallacy. In William W. Cohen and Samuel Gosling, editors, *ICWSM*. The AAAI Press, 2010.
- [5] K.L. Clarkson. Nearest neighbor queries in metric spaces. *Discrete & Computational Geometry*, 22(1):63–93, 1999.
- [6] Alex Fabrikant, Ankur Luthra, Elitza Maneva, Christos H. Papadimitriou, and Scott Shenker. On a network creation game. In *Proc. ACM PODC*, pages 347–351, 2003.
- [7] Alex Fabrikant, Christos H. Papadimitriou, and Kunal Talwar. The complexity of pure nash equilibria. In László Babai, editor, *STOC*, pages 604–612. ACM, 2004.
- [8] Pierre Fraigniaud, Emmanuelle Lebhar, and Laurent Viennot. The inframetric model for the internet. In *Proceedings of the 27th IEEE International Conference on Computer Communications (INFOCOM)*, pages 1085–1093, Phoenix, 2008.
- [9] Anh-Tuan Gai, Dmitry Lebedev, Fabien Mathieu, Fabien De Montgolfier, Julien Reynier, and Laurent Viennot. Acyclic Preference Systems in P2P Networks. In *Proc. Euro-Par*, 2007.
- [10] Sarel Har-Peled and Manor Mendel. Fast construction of nets in low dimensional metrics, and their applications. In Joseph S. B. Mitchell and Günter Rote, editors, *Symposium on Computational Geometry*, pages 150–158. ACM, 2005.
- [11] M.O. Jackson. *Social and Economic Networks*. Princeton University Press. Princeton University Press, 2010.
- [12] David R. Karger and Matthias Ruhl. Finding nearest neighbors in growth-restricted metrics. In John H. Reif, editor, *STOC*, pages 741–750. ACM, 2002.
- [13] Nikolaos Laoutaris, Laura J. Poplawski, Rajmohan Rajaraman, Ravi Sundaram, and Shang-Hua 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.
- [14] Laurent Massoulié and Andy Twigg. Rate-optimal schemes for peer-to-peer live streaming. *Journal of Performance Analysis*, 2008.
- [15] D Monderer and LS Shapley. Potential games. *Games and Economic Behavior*, pages 124–143, 1996.
- [16] Noam Nisan, Tim Roughgarden, Éva Tardos, and Vijay V. Vazirani, editors. *Algorithmic Game Theory*. Cambridge Univ Press, 2007.
- [17] C. Greg Plaxton, Rajmohan Rajaraman, and Andréa W. Richa. Accessing nearby copies of replicated objects in a distributed environment. *Theory Comput. Syst.*, 32(3):241–280, 1999.
- [18] R. W. Rosenthal. A class of games possessing pure-strategy Nash equilibria. *International Journal of Game Theory*, 2:65–67, 1973.

- [19] Tim Roughgarden. Potential functions and the inefficiency of equilibria. *Proceedings of the International Congress of Mathematicians (ICM)*, 3:1071–1094, 2006.
- [20] Herbert A. Simon. Designing organizations for an information rich world. In Martin Greenberger, editor, *Computers, communications, and the public interest*, pages 37–72. Baltimore, 1971.