The Limits of Popularity-Based Recommendations, and the Role of Social Ties

Marco Bressan[*][†]
Sapienza University of Rome
Rome, Italy

Stefano Leucci[*]
Sapienza University of Rome
Rome, Italy

Alessandro Panconesi[*]
Sapienza University of Rome
Rome, Italy

Prabhakar Raghavan
Google
Mountain View, CA

Erisa Terolli[*]
Sapienza University of Rome
Rome, Italy

ABSTRACT

In this paper we introduce a mathematical model that captures some of the salient features of recommender systems that are based on popularity and that try to exploit social ties among the users. We show that, under very general conditions, the market always converges to a steady state, for which we are able to give an explicit form. Thanks to this we can tell rather precisely how much a market is altered by a recommendation system, and determine the power of users to influence others. Our theoretical results are complemented by experiments with real world social networks showing that social graphs prevent large market distortions in spite of the presence of highly influential users.

1. INTRODUCTION

Recommender systems (RS's) are a paradigmatic example of the interaction between humans and algorithms in the cultural arena. From YouTube videos to books on Amazon and movies on iTunes or Netflix, online choices are increasingly mediated by such algorithms. RS's are popular for their ability to conjure up non-trivial relationships between products or news stories, and as such they are becoming powerful economic actors. It is this economic dimension of RS's that we intend to explore in this paper.

A fundamental question in this context, and the one we intend to tackle in this paper, is the following: to what extent can a market be altered by a RS? For instance, assume that an online bookstore starts adopting a RS. Will unknown books become hits and vice-versa? How and how much will the habit of the typical reader change? A large body of work has investigated questions of this kind at the individual, user level [8, 25, 26, 35]. For instance, a nice experiment described

in [25] suggests that even a simple type of feedback, such as providing the ranking based on the number of downloads, may significantly boost market shares. Studies such as these provide precious insights but it is difficult to derive from them quantitatively accurate predictions about markets in the long-run. Also, it is entirely conceivable that the amplification effects observed be temporary, or not transferable to markets whose sheer size and complexity dwarfs that of artificial settings. Other studies have tried to draw conclusions from behavioural data of real markets [11, ?, 10, 23], but the paucity of available data makes them tentative.

In this paper we try a different approach by introducing a natural model for markets that are governed by a RS. Our model is simple enough to allow for a precise mathematical analysis of long-term behaviours, while at the same time it captures some of the relevant features of RS's. Specifically, there are two ingredients that are known to play an important role in RS's and that we want our model to represent. The first is popularity, *i.e.* the degree of success of an item in its market. Several studies indicate that popularity feedback (*e.g.* number of downloads, user ratings, number of views, etc) can be a powerful determinant of online behaviour [25, 5, 35]. The second ingredient is, for lack of a better terminology, the "web of kinship" among users. Social ties are the paradigmatic example here. They are recognized to be an important factor that shape user choices, and recommenders try to leverage them in a variety of ways [22, 31, 33, 6]. More generally RS's try to exploit different types of similarities between users, such as explicit links in an online social network, similarity between patterns of consumer choice, or even the result of complex computations like matrix factorization. All these diverse situations can be simply and usefully represented, as we show in this paper, by assuming the existence of a (weighted) network connecting the users.

In simplified form, our mathematical model of market evolution is as follows. In the market we have users (or buyers), connected via a network, and products. When a user buys a product, it either follows its personal inclinations (mode led by a private probability distribution over the products) or it follows a recommendation with a certain probability. When it does, it consults a random neighbour in the network, the *recommender node*. The recommender node picks a random product from its own list of products purchased so far, with a probability that is proportional to the popularity of the product in that moment, and the item so selected is bought by the current buyer.

The main theoretical result in this paper is that this type of system, encompassing a large class of recommenders, always converges to a unique, stationary limit. The result holds under very general conditions. First, users can buy products at different rates. User u, say, can be ten times faster than user v and one thousand times slower than user w etc. Second, the probability of following the recommendation may change from person to person – each buyer can have his/her own personal level of trust in recommendations, and also trust his/her friends to differing degrees. Third, the probability with which a product is recommended does not need to be uniform, but it can depend on time. For instance, we could recommend only the last ten items bought, or skew the probability with which an item is recommended in many other ways. Finally, the graph can be any directed graph (so u could follow v's recommendations but not vice versa) and users can have their own personal history of purchases before the recommendation mechanism starts operating. None of these affects the result– the market always converges to a unique limit.

It is natural to think of the network connecting the users as a social network, but it can in fact be used to model many more scenarios. For instance, we can make the popularity of an item directly proportional to its number of sales in the whole market by assuming the network to be a clique, or we can represent a powerful super-user that influences the entire network without being affected by it by assuming the network to be an oriented star.

A nice feature of our result is that the limit is given explicitly, in closed form. This makes it possible to explore important properties of the steady state both analytically and computationally. For instance, we can answer easily our motivating question– can the recommendation mechanism alter the market in a significant way? We can easily tell the distance, in whatever norm, between the initial market shares of every product and those at steady state. We can gain insight into what type of topologies amplify or dampen distorting effects, and we can also carry out a fine grained analysis that tells, for each user, how much it affects the final outcome. A by-product of our analysis is a quantitative notion of user influence that pinpoints exactly how much each user affects the market at the limit. Interestingly, this notion turns out to be a generalization of Personalized PageRank (PPR) and it is related to the Shapley values of a coalition game that can be defined in our model.

Our limit theorems provide a rich landscape for computational explorations with real-world social networks. We explore the distorting effects of various graph topologies and the extent to which a user can influence the market. What emerges is a rather nuanced landscape that is spelled out in full in the rest of the paper. Perhaps the most interesting outcome is that real social networks prove themselves to be a bulwark against market distortion, in spite of the presence of celebrities in their midst.

To summarize, this paper is a step in the direction of a systematic investigation of the long-term effects of recommender systems on the markets in which they are operating. This is an important and fascinating topic that deserves more attention that it has gotten so far. We hope that this paper provides some actionable insight into this important problem.

The paper is organized as follows. In § 2 we discuss the relevant literature, in § 3 we define our model and analyse its properties mathematically. We conclude with experiments in § 4.

2. RELATED WORK

Product popularity is known to have powerful feedback effects on user behaviour and choices, and is leveraged by many existing RS's. In one of the most eloquent demonstrations, [25] shows that, in a digital cultural market, just reporting the ranking based on the number of downloads significantly affects user choices and increases the market's inequality and unpredictability. Another telling illustration, [35], proves that users are likely to change their mind and reverse their previous choices when given information about the popularity of items. A study of friend recommendation in Twitter, [28], shows that recommendations of popular users are much more likely to be accepted than recommendations of "average" users. Several existing online marketplaces make use of popularity-based recommendations – for instance Amazon[1], Yelp[2], TripAdvisor[3], SoundCloud[4] and Last.fm[5] all suggest items that are currently trending.

Social ties are also understood to play a key role in the recommendation and purchase processes, and are increasingly exploited by modern RS's. A study on the Epinions[6] network [7] finds that a large part of the items adopted by a user are also adopted by his/her friends. Similarly, a study on the Yahoo! Pulse network [30] shows that a user's interests are highly correlated with, and can be predicted by, the interests of his/her social ties – and vice versa. An emblematic example of RS based on social ties is [31], which uses a generative influence model where users probabilistically pick items according to both their and their friends' preferences. Another study, [6], builds a recommendation algorithm assuming that the two forces driving a user's actions are latent preferences and the influence of friends. Several other works show or propose that recommendations take place over some social network structure [16, 17, 33, 15, 34].

On the other hand, investigations on the long-term market effects of RS's are scarce; as long as mathematical models are concerned, existing work seems confined to the economic literature. To the best of our knowledge, the only theoretical investigation on the market effects of RS's is [12]. In their work, a single user repeatedly buys one of two products according to either his/her personal taste or to a recommendation consisting in the most-bought product so far; depending on the choice of parameters, the process can lead to a concentration of sales. There are many differences between their model and ours – in particular, our model allows users to influence each other. Two earlier studies, [9] on consensus reaching and [13] on opinion formation, investigated the convergence of personal beliefs of users interacting in a social group, similarly to how we investigate convergence in a market with social ties. Their models are however far from ours in many aspects; for instance, ours is stochastic and allows for complex dependences on past events. Another paper, [14], studies the effects of word-of-mouth on product sales, but focusing on equilibrium strategies for maximizing

---

[1] http://www.amazon.com

[2] http://www.yelp.com

[3] http://www.tripadvisor.com

[4] http://www.soundcloud.com

[5] http://www.last.fm

[6] http://www.epinions.com

user satisfaction. Finally, a few other works have analysed the interactions between RS's and markets, but either with different aims [18, 24] or by simulation instead of analysis [32].

3. A MODEL FOR RECOMMENDER SYSTEMS

Let us begin with the formal definition of our model.

Products.

We have a set $\mathcal{P}$ of m products; each product can be bought multiple times during the purchasing process. Products are bought one at a time in an infinite sequence of time steps $t = 1, 2, \ldots$. The purchasing process is specified below.

Users (or Buyers).

We have a set $\mathcal{U} := \{1, \ldots, n\}$ of n users. To each user u is associated:

- a fixed purchasing rate $f_u \in (0, 1)$ representing the rate at which u buys products. The f_u's form a probability distribution, i.e. $\sum_{u \in \mathcal{U}} f_u = 1$. The precise role of f_u is clarified below when the purchasing process is defined

- a probability distribution B_u over the set of products $\mathcal{P}$ reflecting its *personal preferences*. The probability $B_u(p), p \in \mathcal{P}$ is u's *personal preference* for p.

- a fixed probability $\alpha_u \in [0, 1)$; when u buys a product it follows a recommendation with probability α_u (see the definition of the purchasing process below)

- a list of products purchased in the past (before time 0), initially containing $k_u > 0$ items. This list can be arbitrary

The graph.

Users are connected via a directed, weighted graph $G = (\mathcal{U}, E)$, possibly with self-loops. Intuitively, an edge vu denotes the fact that v can influence u when u decides to follow a recommendation, or alternatively, that u "trusts" v. The arc has weight w_{vu} which gives the strength of this trust relationship. The weights are normalised: the weights of the arcs entering each vertex u (if any) sum up to one.

It is natural to think of this graph as a social network among the users, but it is in fact a very versatile tool with which we can model several different situations. For instance, when the graph is a clique and the weights are uniform we have the traditional popularity feedback – an item is recommended with a probability that is proportional to the number of sales in the whole market. Or, the graph could embed similarities between users, with the weights reflecting the degree of similarity, and so on so forth. The graph is a flexible tool thanks to which several situation can be cast in our framework and network effects be explored.

The purchasing process.

At each time step $t = 1, 2, \ldots$, a user u is chosen with probability f_u and buys a product in one of two ways:

- With probability $1 - \alpha_u$, it chooses a product at random according to its own distribution B_u of personal preferences.

- Alternatively, with probability α_u, it follows the recommendation:

  - First, u picks a *recommender node* v at random among those pointing to him (if u has no entering arcs then $\alpha_u = 0$). The recommender node v is picked with probability w_{vu}, the weight of the arc vu.

  - Once the recommender node v is chosen, a product p is selected at random from the multiset of products acquired by v up to that moment. Let us denote for now this probability simply as $x_v^t(p)$ and postpone to later its precise definition.

  - The product p so selected is bought, i.e. it is added to the multiset of products purchased by the current buyer u.

Note that even if we say for convenience that the i-th purchase takes place at "time" i, we make no assumptions about time itself (i.e. the wall clock time between any one purchase and the next can be arbitrary). The process simply stipulates that items are bought one after the other, in an infinite sequence.

The weight of history.

To conclude the description of the model we need to define the probability distribution, that we denoted as $\{x_v^t(p)\}_{p \in \mathcal{P}}$, with which the products of the recommender node are picked. A reasonable choice to model popularity-based recommendations would be simply to select a product uniformly at random from v's multiset of past purchases. In this way popular items have a higher probability of being recommended. But we want to take a more general approach by allowing time-dependent weights. For instance, we would like to be able to deal with a situation in which, say, only the ten most recent purchases matter, or where the influence of purchases fades away with time in a specific way, and so on so forth. To this end, focus on a vertex u and recall that we allow users to have made k_u purchases before starting the system. Suppose now that we are at time $t > 0$, and let $I_u^t \subseteq [t-1]$ denote the set of times when u previously made a purchase. We denote the weight of a purchase made by u at time $i < t$ as $h_u^{t,i}$. First, we assume these weights to be normalised, that is, for all $t > 0$,

$$\sum_{i=1}^{k_u} h_u^{t,-i} + \sum_{i \in I_u^t} h_u^{t,i} = 1,$$

For a given product p, let $J_{u,p}^t \subseteq \{-k_u, \ldots, -1\} \cup I_u^t$ be the set of times when u bought p. Then, given that u is the recommender node at time t, product p will be recommended by u and purchased by the buyer with probability

$$x_u^t(p) := \sum_{i \in J_{u,p}^t} h_u^{t,i}.$$

This formalization captures a large spectrum of natural scenarios. For example, if the recommendation consists in picking from u's past purchases uniformly at random, this amounts to setting $h_u^{t,-i} = h_u^{t,i} = 1/k_u + |I_u^t|$ for all i. Picking one at random among the 10 most recent purchases corresponds instead to setting $h_u^{t,i} = 1/10$ for all i such that $|I_u^i| \geq |I_u^t| - 10$ and $h_u^{t,i} = 0$ otherwise. Note that the values of the weights $h_u^{t,i}$ depend on the outcome of the process, thus the $h_u^{t,i}$ are random variables themselves.

While our theorems will be given for the full model, it may help the intuition to consider the following simplified version, to which we will sometime refer in the remainder of the paper.

DEFINITION 1. [THE BASIC SCENARIO] *The special case where users buy at the same rate (i.e. $f_u = 1/n, u \in \mathcal{U}$), follow recommendations with the same probability (i.e. $\alpha_u = \alpha, u \in \mathcal{U}$), pick the recommender with uniform probability (i.e. $w_{vu} = 1/\text{indeg}(u), u \in \mathcal{U}$), and recommend items by picking a random product from their list with uniform probability (i.e. $h_u^{t,-i} = h_u^{t,i} = 1/k_u + |I_u^t|, u \in \mathcal{U}$), is called the* basic scenario.

Remarks. Our model only considers products that can be bought multiple times. Food items are a natural example, but there are very many cultural products that fall into this category: going out to a specific club, attending sport events, or participating to concerts at specific venues are just a few examples. The case of items for which multiple purchases make little sense, such as books or movies, is also of great interest but is left for future work. It is worth noting that our model can capture to a certain extent *repeat consumption*, the empirical principle that people tend to re-purchase a product they recently bought [1, 2]. More precisely, our model can mimic the one in [1], where a user buys a product that is either new or drawn from a distribution over his/her purchasing history. We can mimic this model by allowing self-loops in the graph: a user can then pick itself as the recommender node and, with properly defined weights, re-purchase a product he bought in the near past.

3.1   Analysis of the model

To analyse the evolution of the process we focus, from now on, on a particular product $p^* \in \mathcal{P}$, and look at all the others as aggregated into a second product.

Let us begin with some necessary notation. Recall that $x_u^t(p)$ is the probability with which p is recommended and bought given that u is the recommender node at time t. To simplify notation, let $x_u^t(p^*)$ be denoted simply as x_u^t. Similarly, let b_u be u's personal preference for p^*, and let $\boldsymbol{b} = (b_1, \ldots, b_n)$. We want to study the evolution of $\boldsymbol{x}^t := (x_1^t, \ldots, x_n^t)$ as t grows. We define $\boldsymbol{A} = \text{diag}(\alpha_1, \ldots, \alpha_n)$ and denote by $\boldsymbol{M}$ the weighted transposed adjacency matrix of G, i.e. $M_{uv} = w_{vu}$ (which is 0 if $vu \notin G$). For any random vector $\boldsymbol{y} = (y_1, y_2, \ldots)$ we denote by $\mathbf{E}[\boldsymbol{y}]$ the vector $(\mathbf{E}[y_1], \mathbf{E}[y_2], \ldots)$ of the expectations of its components.

With the notation pinned down, we can now address our main question: does the market share of p^* converge? The theorem below states that this happens if the weights $h_u^{t,i}$ satisfy the following condition.

The past fades away, i.e. $\lim_{t \to \infty} \mathbf{E}[h_u^{t,i}] = 0.$

This condition is very natural. It states that past purchases will eventually be forgotten. For instance it is satisfied when the recommended product is chosen uniformly at random or when only the last few purchases have non-zero weight. We can now give our main result, which is proven in the appendix.

THEOREM 1. *If the past fades away, then*

$$\lim_{t \to \infty} \mathbf{E}[\boldsymbol{x}^t] = \boldsymbol{x}^\infty \qquad (1)$$

where $\boldsymbol{x}^\infty = (\boldsymbol{I} - \boldsymbol{AM})^{-1}(\boldsymbol{I} - \boldsymbol{A})\boldsymbol{b}$. Otherwise $\mathbf{E}[\boldsymbol{x}^t]$ might not converge, or converge but not to $\boldsymbol{x}^\infty$.

Let us discuss a few interesting consequences of this result.

First, notice that since the network and the values $\boldsymbol{\alpha} = (\alpha_1, \ldots, \alpha_n)$, $\boldsymbol{f} = (f_1, \ldots, f_n)$, and $\boldsymbol{b}$, stay put and only $\boldsymbol{x}^t$ changes, the theorem gives a complete characterization of the market at the limit. The limit vector $\boldsymbol{x}^\infty$ has a natural and quite useful interpretation, given in the next corollary.

DEFINITION 2. *The fraction of times that a user u has bought p^* at time $t \in \mathbb{N}^+ \cup \{\infty\}$ is called the* local market share *(LMS) of p^* w.r.t. u at time t.*

COROLLARY 1. *For every user u, x_u^∞ is the local market share of p^* w.r.t. u at the limit.*

The proof of this corollary in the general case is slightly technical but in the basic scenario (see Definition 1) follows immediately from Theorem 1. In this scenario, by definition of the model, x_u^t is the fraction of times that u has bought p^* up to time t, *i.e.* it is p^*'s LMS w.r.t. u, and Theorem 1 states that this value converges to x_u^∞.

Consider now the matrix

$$\boldsymbol{L} := (\boldsymbol{I} - \boldsymbol{AM})^{-1}(\boldsymbol{I} - \boldsymbol{A}) \qquad (2)$$

This matrix contains a lot of information about the steady state. Now, since $\boldsymbol{x}^\infty = \boldsymbol{Lb}$ we have that $x_u^\infty = \sum_v L_{uv} b_v$. We can think of x_u^∞ as consisting of a set of "slices", each of size $L_{uv} b_v$, $v \in \mathcal{U}$. Each slice is the contribution of v to x_u^∞ which is, as we have seen, p^*'s LMS w.r.t u at the limit. We thus have a measure of how much each user can influence every other user.

If L_{vu} is how much u influences v then $\sum_v L_{vu}$ quantifies u's influence on the entire market. We can collect these individual market influences in an *influence vector*:

$$\boldsymbol{\gamma}^T := \boldsymbol{f}^T \boldsymbol{L} \qquad (3)$$

where $\boldsymbol{f} = [f_1, \ldots, f_n]$. The vector $\boldsymbol{\gamma}$ is stochastic (its entries sum up to one), and its element γ_u is the relative weight of u's personal distribution in the overall sales distribution at equilibrium. The larger the value of γ_u the greater the influence of u on the final market share of p^*. We thus define the *maximum influence* of any user as,

$$\gamma(G) := \max_{u \in G} \gamma_u = \|\boldsymbol{\gamma}\|_\infty \qquad (4)$$

For instance, in the basic scenario, when G is a star $\gamma_u = \Theta(1)$ if u is the centre and $\gamma_u = \Theta(1/n)$ otherwise, so that $\gamma(G) = \Theta(1)$, while when G is a clique $\gamma(G) = \gamma_u = 1/n$ for every u.

Thanks to Theorem 1 we can now easily answer our motivating question– how much can the market be distorted? By combining $\boldsymbol{\gamma}$ and the vector $\boldsymbol{b}$ of personal preferences we can quantify how much the recommender system distorts (i.e. increases or decreases) the overall market share of p^*.

We define the *market distortion* as:

$$\Delta := \frac{\boldsymbol{f} \cdot \boldsymbol{x}^\infty}{\boldsymbol{f} \cdot \boldsymbol{b}} = \frac{\boldsymbol{\gamma} \cdot \boldsymbol{b}}{\boldsymbol{f} \cdot \boldsymbol{b}}, \qquad (5)$$

where $\cdot$ denotes the dot product. This is the ratio between the market share of p^* at the limit and that without the recommender.

Influence, PageRank and Shapley Values.

The influence vector $\boldsymbol{\gamma}$ has a natural interpretation in terms of personalized PageRank. Consider the graph G^T

obtained by transposing G, the graph among the users, and suppose it has no dangling nodes (i.e. no node with outdegree zero). Let $\boldsymbol{W}$ be the normalised adjacency matrix of G^T, i.e. $W_{uv} = 1/\text{outdeg}(u)$ with the outdegree measured in G^T. Finally, pick any $\alpha \in [0, 1)$ and any n-entry stochastic vector $\boldsymbol{r}$. The personalized PageRank vector $\boldsymbol{p}$ with respect to $\boldsymbol{r}$ with damping factor α on G^T is defined as (see e.g. [19]):

$$\boldsymbol{p}^T = \boldsymbol{r}^T (\boldsymbol{I} - \alpha \boldsymbol{W})^{-1} (1 - \alpha) \tag{6}$$

Compare the above expression with that of $\boldsymbol{\gamma}^T$ given by Equation 3 and the definition of $\boldsymbol{L}$:

$$\boldsymbol{\gamma}^T = \boldsymbol{f}^T (\boldsymbol{I} - \boldsymbol{AM})^{-1} (\boldsymbol{I} - \boldsymbol{A}) \tag{7}$$

The vector $\boldsymbol{\gamma}^T$ is the perfect analogue of $\boldsymbol{p}^T$. Indeed, the two coincide by choosing $\boldsymbol{r} = \boldsymbol{f}$ if each node u in G has $\alpha_u = \alpha$ and has a nonempty set of incoming arcs all with the same weight. Our notion of influence is thus a generalization of the standard personalized PageRank, allowing for personalized values of α, graphs with dangling nodes, and arbitrary arc weights. In fact, our model offers an intuitive interpretation similar to that of PageRank: the influence of nodes can be seen as the stationary distribution of a "reverse" random walk on G which at each step, from the current node u, either moves to an incoming neighbour v with probability $\alpha_u w_{vu}$ or, with probability $1 - \alpha_u$, moves to a random node in G chosen according to the distribution $\{f_v\}_{v \in G}$. This walk can be seen as propagating the "credit" for the choice of a purchased product back to the incoming neighbours, the neighbours of the incoming neighbours and so on, recursively. The purchasing rates play the intuitive role of the personalization vector: the more a user buys, the more the users who influence him are influential overall (and they are indeed reached more often by the random walk).

The influence vector $\boldsymbol{\gamma}$ also has an interpretation in terms of the classical notion of Shapley value, which is a mechanism to determine payoffs in coalition games. Shapley values have been proposed as a mechanism to determine fair compensation for user's participation to recommendation systems [18]. Let b_u denote, as usual, u's personal preference for p^*. Consider now the following game: each user can either participate to the market with b_u or set its preference to 0; the value assigned to the coalition is the total market share of p^*. Then, the following holds.

FACT 1. *The Shapley value of each user u in the above game is exactly $\gamma_u b_u$.*

It is worth pointing out that, uncharacteristically, this Shapley value can be computed in polynomial-time.

Speed of convergence.

Theorem 1 does not tell how much time it takes for the market to reach the steady-state; in fact, this depends on how fast the weights decay with t. In the case of the simple uniform random recommendation, one can prove that the system converges to the steady state reasonably fast:

THEOREM 2. *If the recommender node recommends an item picked at random from its purchasing history, then*

$$\|\mathbf{E}[\mathbf{x}^t] - \boldsymbol{x}^\infty\| = O(t^{\alpha - 1}) \tag{8}$$

where α is the maximum α_u over all users.

The proof is omitted from this extended abstract.

Computing $\boldsymbol{\gamma}$.

Clearly, computing $\boldsymbol{\gamma}$ efficiently is a crucial task to the aims of the present paper. Doing this directly through Equations 2 and 3 requires to explicitly invert the matrix $(\boldsymbol{I} - \boldsymbol{AM})$, which can be prohibitively expensive is the inverse is dense (which is the case if G is well-connected). We instead follow a standard alternative approach. Rewrite $(\boldsymbol{I} - \boldsymbol{AM})$ as $\sum_{i=0}^{+\infty} (\boldsymbol{AM})^i$, plug it into Equation 3 and obtain:

$$\boldsymbol{\gamma}^T = \boldsymbol{f}^T \sum_{i=0}^{+\infty} (\boldsymbol{AM})^i (\boldsymbol{I} - \boldsymbol{A}) \tag{9}$$

Now, if G is sparse, so is $\boldsymbol{M}$ and thus all the terms of the summation in Equation 9. We can then approximate $\boldsymbol{\gamma}$ efficiently by truncating the series, which only requires to keep trace of a sparse matrix and vectors. To be more precise, computing all entries of $\boldsymbol{\gamma}$ up to an additive error $O(\epsilon)$ requires time $O\big((|V| + |E|) \log(\frac{1}{\epsilon}) / \log(\frac{1}{\alpha})\big)$ and space $O(|V| + |E|)$, where $\alpha = \max_{u \in G} \alpha_u$. We use this technique throughout our experiments.

4. EXPERIMENTAL RESULTS

In this section we discuss a series of experiments where the underlying graph of our model is a real world social network.

Theorem 1 says that $\boldsymbol{x}^t$ converges to $\boldsymbol{x}^\infty$ in expectation. A first set of experiments shows that the expectation does captures what happens at the limit, *i.e.* the variance is small. We then investigate the question of market distortion – for a given product p^*, how much does its market share change at the limit? A related question concerns so-called influencers. What is their influence and how much can they change the market? We show that when the underlying network is a real world social network market distortion is negligible and even "celebrities" cannot have significant impact. Vice versa, there are other realistic situations, such as the presence of a super-node acting as a "media tycoon", where certain nodes can conquer a significant share of the market thanks to the recommendation mechanism. We also give a rule of thumb to identify top influencers on the basis of their 2-hop neighbourhood. Finally, we explore recommenders not covered by our theoretical results, in which products are recommended with super-linear probabilities. We find that the results vary widely from graph to graph, and depend also heavily on the starting share of products.

Let us now describe the experimental setup.

4.1 Experimental setup

We run our experiments on six following real-world graphs.

Google+ The ego-Gplus graph from SNAP Datasets [21]. It contains $107,614$ nodes (users) and $13,673,453$ arcs representing a follower-to-followed relationship in the Google+ social network.

Twitter SNAP The ego-Twitter graph from SNAP Datasets. It contains $81,306$ nodes and $2,420,766$ follower-to-followed arcs crawled from Twitter.

Twitter LAW The twitter-2010 graph by the Laboratory for Web Algorithmics, Univ. Milan [4, 3]. It contains $41,652,230$ nodes and $1,468,365,182$ follower-to-followed arcs from Twitter.

Slashdot The soc-sign-Slashdot090221 graph from SNAP Datasets. It contains $82,144$ users and $549,202$ signed

(i.e. "friends" or "foes") arcs between users. We kept only the $425,072$ "friend" arcs.

Yelp A social graph extracted from friendships in the Yelp Dataset Challenge dataset.[7] The graph is undirected and contains $365,759$ users and $1,288,031$ edges.[8]

Facebook The graph from the "List of links" Facebook dataset by MPI-SWS [29].[9] The graph is undirected and contains $63,731$ users and $817,090$ edges.

When the graphs are directed we reverse the edges (except for **Twitter LAW** that was already oriented correctly). This is because in social graphs an edge uv denotes the fact that u "trusts" or "follows" v, while in our model the meaning is that u influences v.

Two kinds of experiments are carried out in the following: simulations and computations. In the first, we simulate the purchasing process for $10,000n$ steps and in the basic scenario (see Definition 1), unless otherwise specified. The probability with which users follow a recommendation is set to $\alpha = 0.2$ for every user. Note that this is a conservative estimate, since the available evidence suggests both the relative influence of friends' recommendations (see e.g. [20, 31]) and the fraction of sales causally imputable to recommender systems (see [27]) to be significantly smaller. We are thus intentionally putting the RS in favourable conditions to distort the market.

In computations, we use the adjacency matrix of the underlying graph and the other parameters of the model to compute explicitly the influence vector $\boldsymbol{\gamma}$ (and related quantities), using the techniques described at the end of Section 3.

Our code was partly written in C++ using Eigen[10] and the GNU Scientific Library[11], and partly in Java using the WebGraph framework[12]. We ran it on a distributed-memory 50-node cluster provided by CINECA[13], where each node is equipped with two ten-core Intel Xeon CPUs and 128GiB of main memory. The source code can be downloaded from `https://github.com/Steven--/recommender`.

4.2 Convergence to the equilibrium

The first experiment checks that the purchasing process unrolls as predicted by our theoretical results. As in § 3.1, we focus on a single product p^*, with the remaining products coalesced into one. Figure 1 shows that $\boldsymbol{x}^t$ quickly converges to $\boldsymbol{x}^\infty$ as t grows (the plot follows a power-law).

Figure 2 shows what happens in the short-term setting. Initially, $\|\boldsymbol{x}_t - \boldsymbol{x}\|_\infty$ decreases rapidly, as expected, but then it starts oscillating around the value 0.22. This behaviour might seem counter-intuitive at first, but it is a natural consequence of the short-memory of the users. As the users only consider the last 100 purchases when making a recommendation, small differences in the number of times p^* is purchased may considerably change the probability with which p^* is recommended. This is a setting where $\mathbf{E}[\boldsymbol{x}^t]$ converges to $\boldsymbol{x}^\infty$ but $\boldsymbol{x}^t$ does not. Note however the following. If we consider the vector $\boldsymbol{z}^t$ containing the local market shares of p^* w.r.t.

---

[7] `https://www.yelp.com/dataset_challenge/`

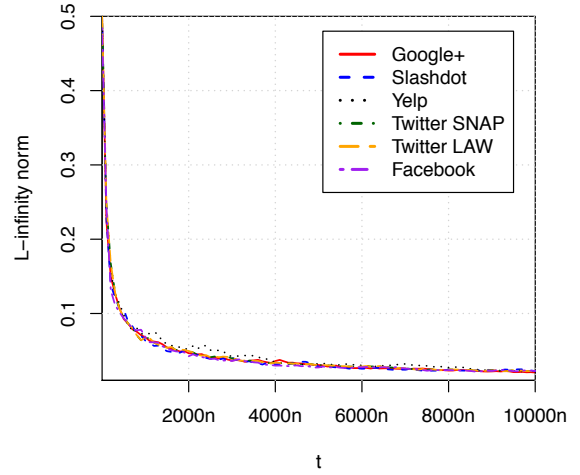[8] Actually, the original Yelp dataset contains 117 spurious directed edges that are not reciprocated.

[9] `http://socialnetworks.mpi-sws.org/data-wosn2009.html`

[10] `http://eigen.tuxfamily.org`

[11] `http://www.gnu.org/software/gsl/`

[12] `http://webgraph.di.unimi.it/`

[13] `http://www.cineca.it/en`

Figure 1: $\|\boldsymbol{x}^t - \boldsymbol{x}^\infty\|_\infty$ **goes to** 0 **as** t **grows (long-memory scenario).**
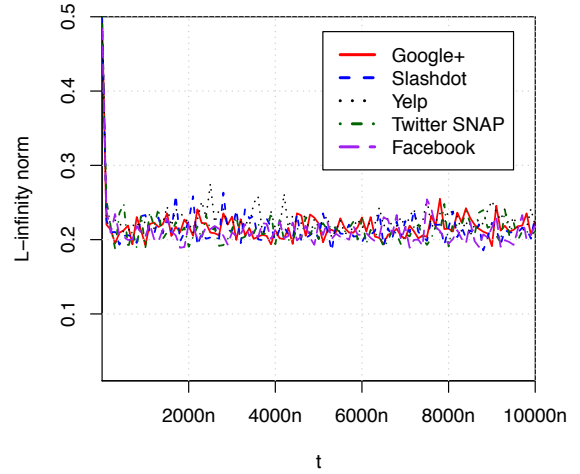

Figure 2: Plot of $\|\boldsymbol{x}^t - \boldsymbol{x}^\infty\|_\infty$ **in the short-memory scenario.**

every user u (Definition 2) then, as predicted by Corollary 1, $\|\boldsymbol{z}^t - \boldsymbol{x}^\infty\|_\infty$ goes to 0 as t increases (see Figure 3). In conclusion, these experiments suggest that the LMS converges not only in expectation but in high probability. We leave this as an open problem.

4.3 Market distortion

In this section we try to understand how much the market can be distorted by the recommendation system. To this end we consider the quantity Δ defined by Equation 5 In the experiments $\boldsymbol{b}$ was instantiated with different types of distributions: uniform in $[0,1]$, exponential of mean $1/2$, a power-law of exponent -0.01 (i.e., $p(x) \propto x^{-0.01}$), and normal of mean $1/2$ and standard deviation $1/6$ (for distributions having support wider than $[0,1]$, the truncated version was used).

In all cases the value of Δ was always in the range 1 ± 0.002, which means that market distortion was negligible for all networks considered. This is an indication that social graphs prevent the recommender from distorting the market.
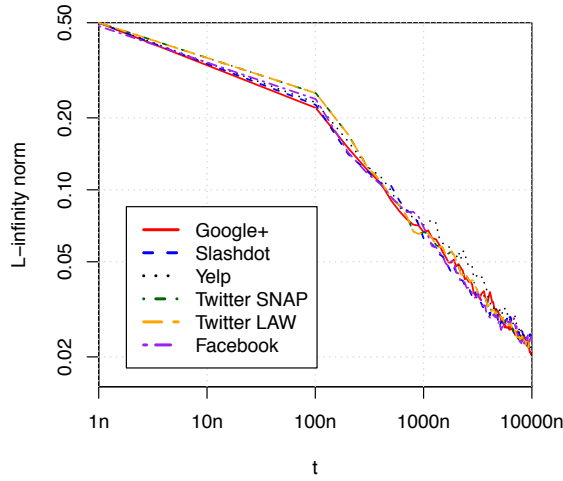
Figure 3: Log-log plot of $\|z^t - x^\infty\|_\infty$ in the short-memory scenario. The vector z^t keeps track, for each user, of the fraction of products p^* bought since the beginning of the process.

The next experiment determines how much (of the overall market share of p^*) can be ascribed to the most influential users. Recall from § 3.1 that $\gamma(G) = \max_u \gamma_u$ is the maximum influence of any user. While $\gamma(G)$ can be seen as a *relative* measure of the importance of the most influential user, i.e., as the fraction of the network that a single user can influence, $n \cdot \gamma(G)$ is an *absolute* measure: the most influential user u can be seen as controlling a group of $n \cdot \gamma(G)$ users that buy p^* with probability b_u at rate $1/n$. Figure 4 shows the values of $n \cdot \gamma(G)$ and of $\gamma(G)$ for the networks we tested. Even if there are users whose purchasing power corresponds to that of more than 80 "average users", none of them can influence a significant portion of the network.

In order to understand if the fairness of these social graphs is an artefact of our model or one of their intrinsic properties, we investigated a modified version of our model not covered by Theorem 1. In the modified model, the probability with which u recommends p^* becomes super-linear and is proportional to x_u^2. Intuitively, this should boost the final market share of p^* as soon as p^* has a small advantage in the user's preference distributions. Therefore, we draw the personal preferences of the users so that the mean of p^* is k times larger than the
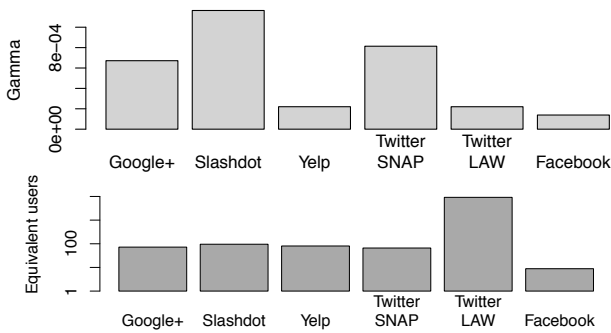
mean of any other product, where the *imbalance k* ranges from 1 to 5. We considered markets having both two and ten different products: the corresponding results can be seen in Figures 6 and 7. The purchasing process was simulated until the change in the market share of the products dropped below 0.001, and the experiment was repeated 10 times in order to compute the standard deviation of Δ, which was always smaller than 0.007. Even with this favourable setting we saw a market distortion effect on social networks that was always less than the one observed on a Clique. We conclude that, although the market share of the leading product p^* can increase significantly when there are many competing products, social graphs do not promote, and often damps, this growth phenomena.

In the last experiment of this section we modify the topology of the networks by inserting a super-node that points to every user and always recommends p^*. This models the effects of mass media such as television, radio, mass advertising, etc. Perhaps unsurprisingly, we observe that such "media tycoons" can influence a relevant portion of the market. Depending on the graph, $\gamma(G)$ ranges from 3.4% to 13.4%. Market distortion can be significant, with values of Δ ranging from 105.3% to 149.7%, as Figure 5 shows.

Figure 5: The effect of media tycoons: variation of the market share of p^* (i.e., Δ) for different graphs and starting distributions when a super-node is inserted.

Figure 4: Fairness of social ties: values of $\gamma(G)$ (light grey) and $n \cdot \gamma(G)$ (dark grey) on different social graphs.
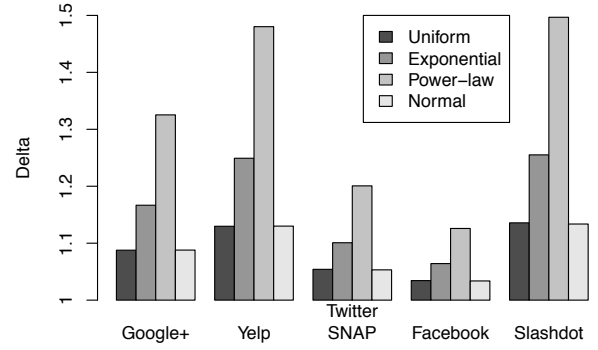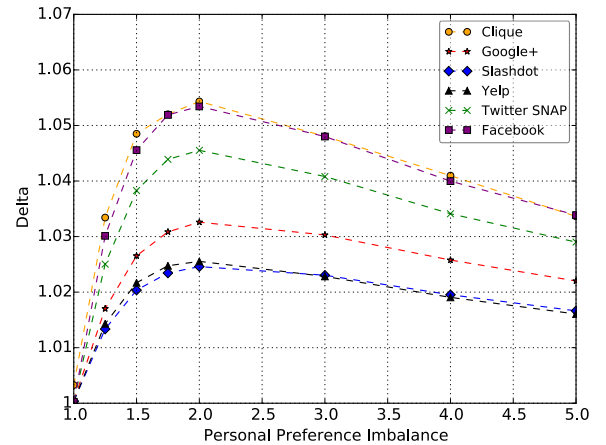

Figure 6: Δ for a market with 2 products as a function of the initial imbalance for product p^*.
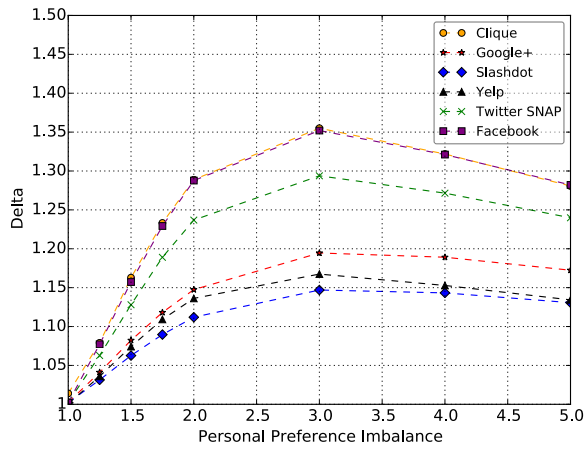
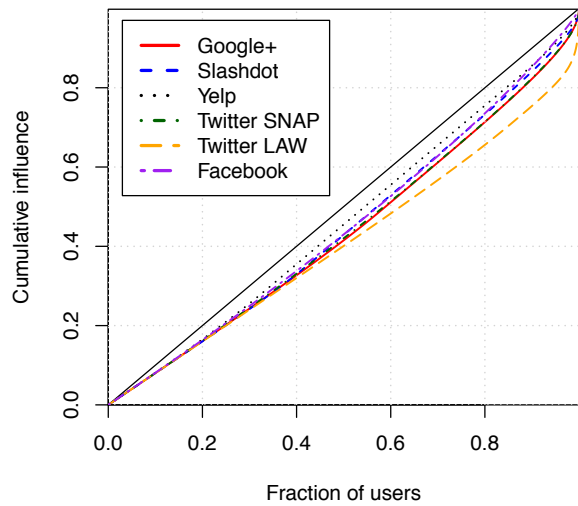Figure 7: Δ for a market with 10 products as a function of the initial imbalance for product p^*.

Figure 8: Lorenz curve of the vector γ for the different graphs. The corresponding Gini indices are 0.128, 0.109, 0.076, 0.125, 0.180, 0.101, in order.

4.4 Looking for the influencers

In this section we look at the distribution of influence in social networks and give a rule of thumb to identify the most influential users.

In a clique every vertex has the same influence ($\gamma_u = 1/n$ for each user u). We want to know how far real social networks are from this perfectly "egalitarian" situation. Figure 8 shows that real social networks are quite fair in the sense that most of the users have an influence that is slightly smaller than $1/n$, while only a small elite of users has a very large influence. The graph where this "oligarchy" is most prominent is `Twitter LAW`, where 10% of the overall influence on the market is due to 0.6% of the users. In all the other networks, at least 4% of all the users are needed to reach the same combined influence.

Next, we try to give a simple rule of thumb to identify the most influential users in a network. The bottom of Figure 9 shows the first one hundred nodes ranked by out-degree from left to right on the x-axis. The y-axis reports their

γ-rank (the shorter the bar the higher the rank). The figure shows that if a node has high out-degree than it will be influential. The top of Figure 9 shows that the converse is not true. Among the one hundred most influential users there are nodes with moderate to low out-degree. Who are these nodes? To illustrate, consider the case of the user whose
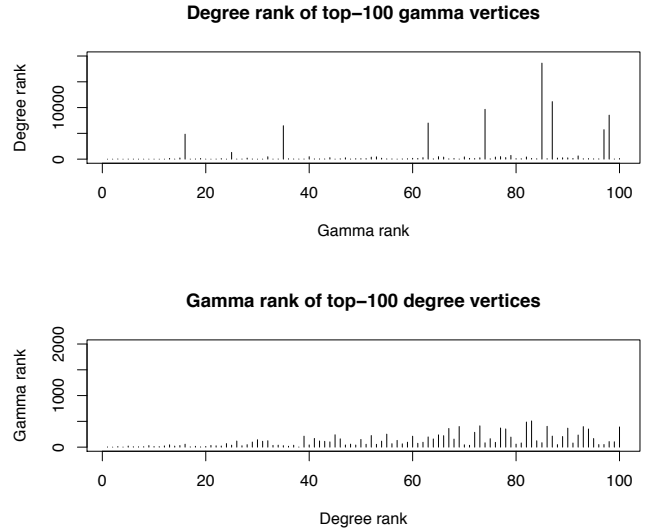

Figure 9: Top: Outdegree rank for the top-100 ranked users w.r.t. γ. Bottom: γ rank for the top-100 ranked users w.r.t. their outdegree. The data refer to `Google+`.

γ-rank in `Google+` is 88. This user lies within the top 1% of the influential users, but his/her 243 followers place him/her in position $11,155$ of the out-degree ranking, outside of the top 10%. This user however has 48 followers each of which follows 3 users or less and, among these, 34 happen to follow only him. In other words, this user is highly "trusted" in a small community. When this happens, the node will have a high γ-rank.

Finally, consider a user u who is trusted by a user v having high degree. We know already that v will be influential. Suppose now that v follows a small number of people that is, u is among the few "confidants" of an influential user. Then, u too will be influential.

To summarize, we can characterize influential users as users who have at least one of the following properties:

- A large number of followers.

- A non-negligible group of followers which, in turn, follow few other users.

- Are among the few friends of an influential user.

This rule of thumb gives a heuristic characterization of all influential users.

5. REFERENCES

[1] A. Anderson, R. Kumar, A. Tomkins, and S. Vassilvitskii. The dynamics of repeat consumption. In *Proc. of WWW*, pages 419–430. 2014.

[2] A. R. Benson, R. Kumar, and A. Tomkins. Modeling User Consumption Sequences. In *Proc. of WWW*, pages 519–529. 2016.

[3] P. Boldi, M. Rosa, M. Santini, and S. Vigna. Layered label propagation: A multiresolution coordinate-free ordering for compressing social networks. In *Proc. of WWW*, pages 587–596. 2011.

[4] P. Boldi and S. Vigna. The WebGraph framework I: Compression techniques. In *Proc. WWW*, pages 595–601. 2004.

[5] O. Celma and P. Cano. From hits to niches?: Or how popular artists can bias music recommendation and discovery. In *Proc. NETFLIX Competition*, pages 5:1–5:8. ACM.

[6] A. J. Chaney, D. M. Blei, and T. Eliassi-Rad. A probabilistic model for using social networks in personalized item recommendation. In *Proc. of ACM RecSys*, pages 43–50. 2015.

[7] F. C. T. Chua, H. W. Lauw, and E.-P. Lim. Predicting Item Adoption Using Social Correlation. In *Proc. of SIAM ICDM*, pages 367–378. 2011.

[8] D. Cosley, S. K. Lam, I. Albert, J. A. Konstan, and J. Riedl. Is seeing believing?: How recommender system interfaces affect users' opinions. In *Proc. of CHI*, pages 585–592. 2003.

[9] M. H. DeGroot. Reaching a consensus. *Journal of the American Statistical Association* 69(345):118 – 121, 1974.

[10] W. Duan, B. Gu, and A. B. Whinston. The dynamics of online word-of-mouth and product sales – an empirical investigation of the movie industry. *Journal of Retailing*, 84(2):233 – 242, 2008.

[11] T. Ehrmann and H. Schmale. The hitchhiker's guide to the long tail: The influence of online-reviews and product recommendations on book sales – evidence from german online retailing. In *Proc. ICIS*, 2008.

[12] D. Fleder and K. Hosanagar. Blockbuster culture's next rise or fall: The impact of recommender systems on sales diversity. *Management science*, 55(5):697–712, 2009.

[13] N. E. Friedkin and E. C. Johnson. Social influence and opinions. *Journal of Mathematical Sociology*, 15(3-4):193–206, 1990.

[14] A. Hervas-Drane. Recommended for you: The effect of word of mouth on sales concentration. *International Journal of Research in Marketing*, 32(2):207–218, 2015.

[15] J. Huang and X. Hu. Information passing in online recommendation. In *Proc. of UEO*, pages 3–6. 2013.

[16] M. Jamali and M. Ester. Trustwalker: A random walk model for combining trust-based and item-based recommendation. In *Proc. of KDD*, pages 397–406. 2009.

[17] M. Jamali and M. Ester. A matrix factorization technique with trust propagation for recommendation in social networks. In *Proc. of ACM RecSys*, pages 135–142. 2010.

[18] J. Kleinberg, P. Raghavan, and C. Papadimitriou. On the value of private information. In *Proc. TARK*, 2001.

[19] A. N. Langville and C. D. Meyer. Deeper inside PageRank. *Internet Mathematics*, 1(3):335–380, 2004.

[20] J. Leskovec, L. A. Adamic, and B. A. Huberman. The dynamics of viral marketing. *ACM Trans. Web*, 1(1), May 2007.

[21] J. Leskovec and A. Krevl. SNAP Datasets: Stanford large network dataset collection. http://snap.stanford.edu/data.

[22] H. Ma, I. King, and M. R. Lyu. Learning to recommend with social trust ensemble. In *Proc. of ACM SIGIR*, pages 203–210. 2009.

[23] T. T. Nguyen, P.-M. Hui, F. M. Harper, L. Terveen, and J. A. Konstan. Exploring the filter bubble: The effect of using recommender systems on content diversity. In *Proc. WWW*, pages 677–686. 2014.

[24] S. Prawesh and B. Padmanabhan. The "Top N" news recommender: Count distortion and manipulation resistance. In *Proc. of ACM RecSys*, pages 237–244. 2011.

[25] M. J. Salganik, P. S. Dodds, and D. J. Watts. Experimental study of inequality and unpredictability in an artificial cultural market. *Science*, 311(5762):854–856, 2006.

[26] A. Sharma and D. Cosley. Do social explanations work?: Studying and modeling the effects of social explanations in recommender systems. In *Proc. of WWW*, pages 1133–1144. 2013.

[27] A. Sharma, J. M. Hofman, and D. J. Watts. Estimating the Causal Impact of Recommendation Systems from Observational Data. In *Proc. of ACM EC*, pages 453–470. 2015.

[28] J. Su, A. Sharma, and S. Goel. The effect of recommendations on network structure. In *Proc. of WWW*. 2016 (to appear).

[29] B. Viswanath, A. Mislove, M. Cha, and K. P. Gummadi. On the evolution of user interaction in Facebook. In *Proc. of ACM WOSN*, pages 37–42. 2009.

[30] S.-H. Yang, B. Long, A. Smola, N. Sadagopan, Z. Zheng, and H. Zha. Like like alike: Joint friendship and interest propagation in social networks. In *Proc. of WWW*, pages 537–546. 2011.

[31] M. Ye, X. Liu, and W.-C. Lee. Exploring social influence for recommendation: A generative model approach. In *Proc. of ACM SIGIR*, pages 671–680. 2012.

[32] A. Zeng, C. H. Yeung, M. Medo, and Y.-C. Zhang. Modeling mutual feedback between users and recommender systems. *Journal of Statistical Mechanics: Theory and Experiment*, 2015(7):P07020, 2015.

[33] T. Zhao, C. Li, M. Li, Q. Ding, and L. Li. Social recommendation incorporating topic mining and social trust analysis. In *Proc. of ACM CIKM*, pages 1643–1648, 2013.

[34] T. Zhao, J. McAuley, and I. King. Leveraging social connections to improve personalized ranking for collaborative filtering. In *Proc. of ACM CIKM*, pages 261–270. 2014.

[35] H. Zhu and B. A. Huberman. To switch or not to switch. Understanding social influence in online choices. *American Behavioral Scientist*, 58(10):1329–1344, 2014.

APPENDIX

Proof of Theorem 1.

We make use of the following lemma:

LEMMA 1. *The matrix $(\boldsymbol{I} - \boldsymbol{AM})$ is nonsingular, and the matrix $(\boldsymbol{I} - \boldsymbol{AM})^{-1}(\boldsymbol{I} - \boldsymbol{A})$ is row stochastic (i.e. is real, non-negative, and each row sums to 1).*

PROOF. Since $\alpha_u < 1$ for all u and $\boldsymbol{M}$ is row-stochastic, then $(\boldsymbol{I} - \boldsymbol{AM})$ is strictly diagonally dominant; hence it is invertible by the Levy-Desplanques theorem. Now $(\boldsymbol{I} - \boldsymbol{AM})^{-1} = \sum_{i \geq 0}(\boldsymbol{AM})^i$ is real and nonnegative, as well as $(\boldsymbol{I} - \boldsymbol{A})$, hence so is their product $(\boldsymbol{I} - \boldsymbol{AM})^{-1}(\boldsymbol{I} - \boldsymbol{A})$. Finally, since $\boldsymbol{M} \cdot \boldsymbol{1} = \boldsymbol{1}$, then also $(\boldsymbol{I} - \boldsymbol{A}) \cdot \boldsymbol{1} = (\boldsymbol{I} - \boldsymbol{AM}) \cdot \boldsymbol{1}$, and thus $(\boldsymbol{I} - \boldsymbol{AM})^{-1}(\boldsymbol{I} - \boldsymbol{A}) \cdot \boldsymbol{1} = (\boldsymbol{I} - \boldsymbol{AM})^{-1}(\boldsymbol{I} - \boldsymbol{AM}) \cdot \boldsymbol{1} = \boldsymbol{1}$. □

Lemma 1 implies that the vector $\boldsymbol{x}^\infty$ exists and has components in $[0, 1]$. Let us now turn to prove the main statement.

"IF" SIDE. This part is divided in three steps: writing a recurrence equation for $\mathbf{E}[x_u^t]$, converting it into a matricial recurrence equation for $\mathbf{E}[\mathbf{x}^t] - \boldsymbol{x}^\infty$, and showing that $|\mathbf{E}[\mathbf{x}^t] - \boldsymbol{x}^\infty|$ tends to the null vector $\boldsymbol{0}$.

A recurrence equation for $\mathbf{E}[x_u^t]$. The proof makes use of two additional random variables: ψ_u^t, which indicates the event that u buys at step t, and χ_u^t, which indicates the

event that u buys p^* at step t. Then x_u^t can be written as:

$$x_u^t = \sum_{i=1}^{k_u} h_u^{t,-i} \chi_u^{-i} + \sum_{i=1}^{t-1} h_u^{t,i} \chi_u^i \tag{10}$$

where the χ_u^{-i} are actually constants since the initial history is fixed. Gather the first summation in a single random variable $h_u^{t,0}$, and take the expectation of both sides:

$$\mathbf{E}[x_u^t] = \mathbf{E}[h_u^{t,0}] + \sum_{i=1}^{t-1} \mathbf{E}[h_u^{t,i} \chi_u^i] \tag{11}$$

Note that $\mathbf{E}[\cdot z] = \mathbf{E}[\cdot | z = 1]\mathbf{E}[z]$ if z is a binary random variable. Also, the weights do not depend on the specific products bought, so $\mathbf{E}[h_u^{t,i} | \chi_u^i = 1] = \mathbf{E}[h_u^{t,i} | \psi_u^i = 1]$. Using these facts, one can easily rewrite $\mathbf{E}[h_u^{t,i} \chi_u^i]$ as $\mathbf{E}[h_u^{t,i} \psi_u^i]\mathbf{E}[\chi_u^i | \psi_u^i = 1]$. We can then restate Equation 11 in the following way:

$$\mathbf{E}[x_u^t] = \mathbf{E}[h_u^{t,0}] + \sum_{i=1}^{t-1} \mathbf{E}[h_u^{t,i} \psi_u^i]\mathbf{E}[\chi_u^i | \psi_u^i = 1] \tag{12}$$

Let us now focus on $\mathbf{E}[\chi_u^i | \psi_u^i = 1]$, the probability that u buys p^* at step i given that he buys. By definition of the purchasing process, we have:

$$\mathbf{E}[\chi_u^i | \psi_u^i = 1] = (1 - \alpha_u)b_u + \alpha_u \sum_{(v,u) \in G} w_{vu} \mathbf{E}[x_v^i | \psi_u^i = 1] \tag{13}$$

However, we can replace $\mathbf{E}[x_v^i | \psi_u^i = 1]$ by $\mathbf{E}[x_v^i]$ as x_v^i only depends on v's past purchases. Plugging the resulting expression for $\mathbf{E}[\chi_u^i | \psi_u^i = 1]$ into Equation 12, we obtain the following recurrence relation:

$$\mathbf{E}[x_u^t] = \mathbf{E}[h_u^{t,0}] + \sum_{i=1}^{t-1} \mathbf{E}[h_u^{t,i} \psi_u^i]((1-\alpha_u)b_u + \alpha_u \sum_{(v,u) \in G} w_{vu} \mathbf{E}[x_v^i]) \tag{14}$$

A recurrence equation for $\mathbf{E}[\mathbf{x}^t] - \boldsymbol{x}^\infty$. We next turn Equation 14 into an expression for $\mathbf{E}[\mathbf{x}^t]$. For all $t > 0$ let $\mathbf{h}_{t,0}$ be the vector $[\mathbf{E}[h_1^{t,0}], \dots, \mathbf{E}[h_n^{t,0}]]$; similarly, for all $t > i > 0$ let $\mathbf{H}_{t,i}$ be the matrix $\mathrm{diag}(\mathbf{E}[h_1^{t,i} \psi_1^i], \dots, \mathbf{E}[h_n^{t,i} \psi_n^i])$. Finally, recall the definitions of $\boldsymbol{A}, \boldsymbol{M}, \boldsymbol{b}, \mathbf{x}^i$. The following matricial counterpart to Equation 14 holds:

$$\mathbf{E}[\mathbf{x}^t] = \mathbf{h}_{t,0} + \sum_{i=1}^{t-1} \mathbf{H}_{t,i}((\boldsymbol{I} - \boldsymbol{A})\boldsymbol{b} + \boldsymbol{A}\boldsymbol{M}\mathbf{E}[\mathbf{x}^i]) \tag{15}$$

We now subtract $\boldsymbol{x}^\infty$ from both sides – but putting the right-hand side in a convenient form. Indeed, by definition of $\mathbf{h}_{t,0}$ and $\mathbf{H}_{t,i}$ and by the fact that $\sum_{i=1}^{k_u} h_u^{t,-i} + \sum_{i=1}^{t-1} h_u^{t,i} \psi_u^i = 1$, if we define $\mathbf{h}_{t,0}'$ as the vector whose u-th component is $\mathbf{E}[h_u^{t,0}] - \mathbf{E}[\sum_{i=1}^{k_u} h_u^{t,-i}]x_u^\infty$, one can write:

$$\mathbf{E}[\mathbf{x}^t] - \boldsymbol{x}^\infty = \mathbf{h}_{t,0}' + \sum_{i=1}^{t-1} \mathbf{H}_{t,i}((\boldsymbol{I} - \boldsymbol{A})\boldsymbol{b} + \boldsymbol{A}\boldsymbol{M}\mathbf{E}[\mathbf{x}^i] - \boldsymbol{x}^\infty) \tag{16}$$

But by definition of $\boldsymbol{x}^\infty$ we have $(\boldsymbol{I} - \boldsymbol{A})\boldsymbol{b} = (\boldsymbol{I} - \boldsymbol{A}\boldsymbol{M})\boldsymbol{x}^\infty$; thus $(\boldsymbol{I} - \boldsymbol{A})\boldsymbol{b} - \boldsymbol{x}^\infty = -\boldsymbol{A}\boldsymbol{M}\boldsymbol{x}^\infty$ and Equation 16 becomes:

$$\mathbf{E}[\mathbf{x}^t] - \boldsymbol{x}^\infty = \mathbf{h}_{t,0}' + \sum_{i=1}^{t-1} \mathbf{H}_{t,i}\boldsymbol{A}\boldsymbol{M}(\mathbf{E}[\mathbf{x}^i] - \boldsymbol{x}^\infty) \tag{17}$$

$|\mathbf{E}[\mathbf{x}^t] - \boldsymbol{x}^\infty|$ **goes to zero.** We finally show that $|\mathbf{E}[\mathbf{x}^t] - \boldsymbol{x}^\infty|$, the vector of the absolute values of $\mathbf{E}[\mathbf{x}^t] - \boldsymbol{x}^\infty$, vanishes

with t. The intuition is that $\mathbf{h}_{t,0}'$ and some of the first terms of the summation in Equation 17 go to zero with t by hypothesis, and the rest is damped by the norm of $\boldsymbol{A}$ which is strictly smaller than 1; this gives a bound on $|\mathbf{E}[\mathbf{x}^t] - \boldsymbol{x}^\infty|$ which can be plugged in again, proving the thesis by induction.

Hereafter, inequalities are meant component-wise. First, since $\mathbf{H}_{t,i}, \boldsymbol{A}, \boldsymbol{M}$ are non-negative, Equation 17 implies:

$$|\mathbf{E}[\mathbf{x}^t] - \boldsymbol{x}^\infty| \le |\mathbf{h}_{t,0}'| + \sum_{i=1}^{t-1} \mathbf{H}_{t,i}\boldsymbol{A}\boldsymbol{M}|\mathbf{E}[\mathbf{x}^i] - \boldsymbol{x}^\infty| \tag{18}$$

Suppose now $|\mathbf{E}[\mathbf{x}^t] - \boldsymbol{x}^\infty| \le c\mathbf{1}$ for all $t > \tau$, for some constants $\tau, c \ge 0$; this trivially holds for $\tau = 0$ and $c = 1$. One can check that, by hypothesis of forgetfulness, each element of $\mathbf{h}_{t,0}'$ goes to zero with t, and the same does each element of $\mathbf{H}_{t,i}$ for any fixed i. Thus for any given $\epsilon > 0$ there exists a $t' \ge \tau$ such that for all $t > t'$:

$$|\mathbf{h}_{t,0}'| + \sum_{i=1}^{\tau} \mathbf{H}_{t,i}\boldsymbol{A}\boldsymbol{M}|\mathbf{E}[\mathbf{x}^i] - \boldsymbol{x}^\infty| \le \epsilon \tag{19}$$

and then Equation 18 gives, for all $t > t'$:

$$|\mathbf{E}[\mathbf{x}^t] - \boldsymbol{x}^\infty| \le \epsilon\mathbf{1} + \sum_{i=\tau+1}^{t-1} \mathbf{H}_{t,i}\boldsymbol{A}\boldsymbol{M}|\mathbf{E}[\mathbf{x}^i] - \boldsymbol{x}^\infty| \tag{20}$$

$$\le \epsilon\mathbf{1} + \sum_{i=\tau+1}^{t-1} \mathbf{H}_{t,i}\boldsymbol{A}\boldsymbol{M} \cdot c\mathbf{1} \tag{21}$$

Each row of $\boldsymbol{A}\boldsymbol{M}$ sums to at most α where $\alpha = \max_{v \in G} \alpha_v$; while, since $\sum_{i=1}^{t-1} h_u^{t,i} \psi_u^i \le 1$, each row of $(\sum_{i=\tau+1}^{t-1} \mathbf{H}_{t,i})$ sums to at most 1. Each row of $\sum_{i=\tau+1}^{t-1} \mathbf{H}_{t,i}\boldsymbol{A}\boldsymbol{M}$ thus sums to at most α. And then for all $t > t'$:

$$|\mathbf{E}[\mathbf{x}^t] - \boldsymbol{x}^\infty| \le \epsilon\mathbf{1} + \alpha c\mathbf{1} \tag{22}$$

which for e.g. $\epsilon = \frac{1-\alpha}{2}c$ gives $|\mathbf{E}[\mathbf{x}^t] - \boldsymbol{x}^\infty| \le \frac{1+\alpha}{2}c\mathbf{1}$. But we can then repeat the whole argument multiple times, each time replacing τ by t' and c by $\frac{1+\alpha}{2}c$ and tightening the bound on $|\mathbf{E}[\mathbf{x}^t] - \boldsymbol{x}^\infty|$ by a factor $\frac{1+\alpha}{2}$ for all sufficiently large t. Since $\alpha < 1$ and thus $\frac{1+\alpha}{2} < 1$, this implies

$$\lim_{t \to +\infty} |\mathbf{E}[\mathbf{x}^t] - \boldsymbol{x}^\infty| = \mathbf{0} \tag{23}$$

which concludes this part of the proof.

"ONLY IF" SIDE. We show a simple counterexample. Pick a G where all u have incoming arcs and let all $\alpha_u > 0$. Also, let all u buy the same product a at time -1 and set $h_u^{t,-1} = 1$ for every t. All nodes thus always recommend a, so if $\lim_{t \to \infty} \mathbf{E}[\mathbf{x}^t]$ exists it must depend on whether or not $p^* = a$; thus it cannot be $\boldsymbol{x}^\infty$. Moreover, one can make the system never converge by putting two different products a and b in the initial purchasing history of all users and making them periodically recommend only a or only b.