Diffusion and Cascading Behavior in Random Networks

Marc Lelarge
INRIA-ENS
Paris, France
marc.lelarge@ens.fr

ABSTRACT

The spread of new ideas, behaviors or technologies has been extensively studied using epidemic models. Here we consider a model of diffusion where the individuals' behavior is the result of a strategic choice. We study a simple coordination game with binary choice and give a condition for a new action to become widespread in a random network. We also analyze the possible equilibria of this game and identify conditions for the coexistence of both strategies in large connected sets. Finally we look at how can firms use social networks to promote their goals with limited information. Our results differ strongly from the one derived with epidemic models and show that connectivity plays an ambiguous role: while it allows the diffusion to spread, when the network is highly connected, the diffusion is also limited by high-degree nodes which are very stable.

Categories and Subject Descriptors

G.3 [**Mathematics of Computing**]: PROBABILITY AND STATISTICS

Keywords

social networks, diffusion, random graphs, empirical distribution

1. INTRODUCTION

To illustrate our point, consider the basic game-theoretic diffusion model proposed in [7]. Consider a graph G in which the nodes are the individuals in the population and there is an edge (i, j) if i and j can interact with each other. Each node has a choice between two possible behaviors labelled A and B. On each edge (i, j), there is an incentive for i and j to have their behaviors match, which is modeled as the following coordination game parameterised by a real number $q \in (0, 1)$: if i and j choose A (resp. B), they each receive a payoff of q (resp. $(1 - q)$); if they choose opposite strategies, then they receive a payoff of 0. Then the total payoff of a player is the sum of the payoffs with each of her neighbors. Consider a network where all nodes initially play A. If a small number of nodes are forced to adopt strategy B and other nodes in the network apply best-response updates, then these nodes will be repeatedly applying the following rule: switch to B if enough of your neighbors have already adopted B. There can be a cascading sequence of nodes switching to B such that a network-wide equilibrium is reached in the limit. Most of the results on this model are restricted to deterministic (possibly infinite) graphs. In this work, we analyze the diffusion in the large population limit when the underlying graph is a random network $G(n, \mathbf{d})$ with n vertices and where $\mathbf{d} = (d_i)_1^n$ is a given degree (i.e. number of neighbors) sequence, similarly to [4].

In this simple model, agents play a local interaction binary game where the underlying social network is modeled by a sparse random graph. First considering the deterministic best response dynamics, we compute the contagion threshold for this model, confirming the heuristic result of [8]. We find that when the social network is sufficiently sparse, the contagion is limited by the low connectivity of the network; when it is sufficiently dense, the contagion is limited by the stability of the high-degree nodes. This phenomenon explains why contagion is possible only in a given range of the global connectivity (i.e. the average number of neighbors).

We identify the set of agents able to trigger a large cascade: the pivotal players, i.e. the largest component of players requiring a single neighbor to change strategy in order to follow the change. When contagion is possible, both in the low and high-connectivity cases, the number of pivotal players is low, resulting in rare occurences of cascades. However in the high-connectivity case, we found that the system displays a robust-yet-fragile quality: while the cascades are very rare, their sizes are very large. This feature makes global contagions exceptionally hard to anticipate.

Motivated by social advertising, we also consider cases where contagion is not possible if the set of initial adopters is too small, i.e. a negligible fraction of the total population, as in [3]. We compute the final size of the contagion as a function of the fraction of the initial adopters. We find that the low and high-connectivity cases still have different features: in the first case, the global connectivity helps the spread of the conatgion while in the second case, high connectivity inhibits the global contagion but once it occurs, it facilitates its spread.

We also analyze possible equilibria of the game and in particular, we find conditions for the existence of equilibria with co-existent conventions. Finally, we analyze a gen-

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, to republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.
MAMA 2011
Copyright 20XX ACM X-XXXXX-XX-X/XX/XX ...$10.00.

eral percolated threshold model for the diffusion allowing to give different weights to the (anonymous) neighbors. This model allows us to study rigorously semi-anonymous threshold games of complements with local interactions on a complex network. Our general analysis gives explicit formulas for the spread of the diffusion in terms of the initial condition, the degree sequence of the random graph, and the distribution of the thresholds.

We refer to [5] for these last two points.

2. ANALYSIS OF A SIMPLE MODEL OF CASCADES

2.1 Graphs: the configuration model

We consider a set $[n] = \{1, \dots, n\}$ of agents interacting over a social network. Let $\boldsymbol{d} = (d_i^{(n)})_1^n = (d_i)_1^n$ be a sequence of non-negative integers such that $\sum_{i=1}^n d_i$ is even. For notational simplicity we will usually not show the dependency on n explicitly. This sequence is the degree sequence of the graph: agent $i \in [n]$ has degree d_i, i.e. has d_i neighbors. We define a random multigraph (allowing for self-loop and multiple links) with given degree sequence $\boldsymbol{d}$, denoted by $G^*(n, \boldsymbol{d})$ by the configuration model [1]. Conditioned on the multigraph $G^*(n, \boldsymbol{d})$ being a simple graph, we obtain a uniformly distributed random graph with the given degree sequence, which we denote by $G(n, \boldsymbol{d})$.

We will let $n \to \infty$ and assume that we are given $\boldsymbol{d} = (d_i)_1^n$ satisfying the following regularity conditions, see [6]:

CONDITION 1. *For each n, $\boldsymbol{d} = (d_i^{(n)})_1^n$ is a sequence of non-negative integers such that $\sum_{i=1}^n d_i$ is even and, for some probability distribution $\boldsymbol{p} = (p_r)_{r=0}^\infty$ independent of n,*

(i) $|\{i : d_i = r\}|/n \to p_r$ *for every* $r \geq 0$ *as* $n \to \infty$;

(ii) $\lambda := \sum_{r \geq 0} r p_r \in (0, \infty)$;

(iii) $\sum_{i \in [n]} d_i^2 = O(n)$.

(iv) $\sum_{i \in [n]} d_i^3 = O(n)$.

In words, we assume that the empirical distribution of the degree sequence converges to a fixed probability distribution $\boldsymbol{p}$ with a finite mean λ.

2.2 Contagion threshold for random networks

An interesting perspective is to understand how different network structures are more or less hospitable to cascades. Going back to previous diffusion model, we see that the lower q is, the easiest the diffusion spreads. In [7], the contagion threshold of a connected infinite network (called the cascade capacity in [2]) is defined as the maximum threshold q_c at which a finite set of initial adopters can cause a complete cascade, i.e. the resulting cascade of adoptions of B eventually causes every node to switch from A to B. There are two possible models to consider depending whether the initial adopters changing from A to B apply or not best-response update. It is shown in [7] that the same contagion threshold arises in both models. In this section, we restrict ourselves to the model where the initial adopters are forced to play B forever. In this case, the diffusion is monotone and the number of nodes playing B is non-decreasing. We say that this case corresponds to the permanent adoption model: a player playing B will never play A again.

We now compute the contagion threshold for a sequence of random networks. Since a random network is finite and not necessarily connected, we first need to adapt the definition of contagion threshold to our context. For a graph $G = (V, E)$ and a parameter q, we consider the largest connected component of the induced subgraph in which we keep only vertices of degree strictly less than q^{-1}. We call the vertices in this component pivotal players: if only one pivotal player switches from A to B then the whole set of pivotal players will eventually switch to B in the permanent adoption model. For a player $v \in V$, we denote by $C(v, q)$ the final number of players B in the permanent adoption model with parameter q, when the initial state consists of only v playing B, all other players playing A. Informally, we say that $C(v, q)$ is the size of the cascade induced by player v.

PROPOSITION 2. *Consider the random graph $G(n, \mathbf{d})$ satisfying Condition 1 with asymptotic degree distribution $\boldsymbol{p} = (p_r)_{r=0}^\infty$, and define q_c by:*

$$q_c(\boldsymbol{p}) = q_c \quad = \quad \sup \left\{ q : \sum_{2 \leq r < q^{-1}} r(r-1)p_r > \sum_{1 \leq r} r p_r \right\}.$$

Let $P^{(n)}$ be the set of pivotal players in $G(n, \mathbf{d})$.

(i) For $q < q_c$, there are constants $0 < \gamma(q, \boldsymbol{p}) \leq s(q, \boldsymbol{p})$ such that w.h.p. $\lim_n \frac{|P^{(n)}|}{n} = \gamma(q, \boldsymbol{p})$ and for any $v \in P^{(n)}$, $\liminf_n \frac{C(v,q)}{n} \geq s(q, \boldsymbol{p})$.

(ii) For $q > q_c$, for an uniformly chosen player v, we have $C(v, q) = o_p(n)$. The same result holds if $o(n)$ players are chosen uniformly at random.

This result is in accordance with the heuristic result of [8] (see in particular the cascade condition Eq. 5 in [8]). and is proved in [5]

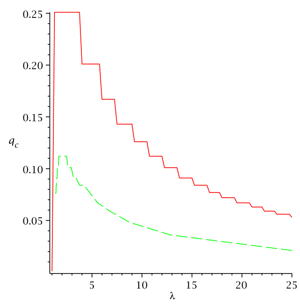

Figure 1: $q_c(\lambda)$ for Erdős-Rényi random graphs and for power law graphs (dashed curve) as a function of the average degree λ.

We now consider some examples. The curves in Figure 1 corresponds to the contagion threshold for an Erdős-Rényi random graphs $G(n, \lambda/n)$ corresponding to $p_r = e^{-\lambda} \frac{\lambda^r}{r!}$ for all $r \geq 0$ and a scale-free random network whose degree distribution $p_r = \frac{r^{-\gamma}}{\zeta(\gamma)}$ (with $\zeta(\gamma) = \sum r^{-\gamma}$) is parameterised by the decay parameter $\gamma > 1$. We see that in this case we have $q_c \leq 1/9$. In other words, in an Erdős-Rényi random graph, in order to have a global cascade, the parameter q must be such that any node with no more than four neighbors must be able to adopt B even if it has a single adopting neighbor. In the case of the scale free random network considered, the parameter q must be much lower and any node with no more than nine neighbors must be able to adopt B with a single adopting neighbor. This simply reflects the intuitive idea that for widespread diffusion to occur there must be a sufficient high frequency of nodes that are certain to propagate the adoption.

We also observe that in both cases, for q sufficiently low, there are two critical values for the parameter λ, $1 < \lambda_i(q) < \lambda_s(q)$ such that a global cascade for a fixed q is only possible for $\lambda \in (\lambda_i(q); \lambda_s(q))$. The heuristic reason for these two thresholds is that a cascade can be prematurely stopped at high-degree nodes. For Erdős-Rényi random graphs, when $1 \leq \lambda < \lambda_i(q)$, there exists a "giant component", i.e. a connected component containing a positive fraction of the nodes. The high-degree nodes are quite infrequent so that the diffusion should spread easily. However, for λ close to one, the diffusion does not branch much and progresses along a very thin tree, "almost a line", so that its progression is stopped as soon as it encounters a high-degree node. Due to the variability of the Poisson distribution, this happens before the diffusion becomes too big for $\lambda < \lambda_i(q)$. Nevertheless the condition $\lambda > \lambda_i(q)$ is not sufficient for a global cascade. Global diffusion also requires that the network not be too highly connected. This is reflected by the existence of the second threshold $\lambda_s(q)$ where a further transition occurs, now in the opposite direction. For $\lambda > \lambda_s(q)$, the diffusion will not reach a positive fraction of the population. The intuition here is clear: the frequency of high-degree nodes is so large that diffusion cannot avoid them and typically stops there since it is unlikely that a high enough fraction of their many neighbors eventually adopts. Following [8], we say that these nodes are locally stable.

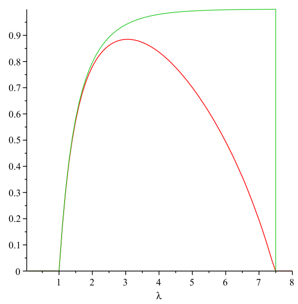

Figure 2: Size $s(q, \lambda)$ of the cascade (in percent of the total population) for Erdős-Rényi random graphs as a function of λ the average degree for a fixed $q = 0.15$. The lower curve gives the asymptotic fraction of pivotal players $\gamma(q, \lambda)$.

The lower curve in Figure 2 represents the number of pivotal players in an Erdős-Rényi random graph as a function of λ the average connectivity for $q^{-1} = 6.666...$. By the same heuristic argument as above, we expect two phase transitions for the size of the set of pivotal players. Indeed the phase transitions occur at the same values $\lambda_i(q)$ and $\lambda_s(q)$ as can be seen on Figure 2 where the normalized size (i.e. fraction) $\gamma(q, \lambda)$ of the set of pivotal players is positive only for $\lambda \in (\lambda_i(q), \lambda_s(q))$. Hence a cascade is possible if and only if there is a 'giant' component of pivotal players. Note also that both phase transitions for the pivotal players are continuous, in the sense that the function $\lambda \mapsto \gamma(q, \lambda)$ is continuous. This is not the case for the second phase transition for the normalized size of the cascade given by $s(q, \lambda)$ in Proposition 2: the function $\lambda \mapsto s(q, \lambda)$ is continuous in $\lambda_i(q)$ but not in $\lambda_s(q)$ as depicted on Figure 2. This has important consequences: around $\lambda_i(q)$ the propagation of cascades is limited by the connectivity of the network as in standard epidemic models. But around $\lambda_s(q)$, the propagation of cascades is not limited by the connectivity but by the high-degree nodes which are locally stable.

2.3 Advertising with word of mouth communication

We consider now scenarios where $\lambda \notin [\lambda_i(q), \lambda_s(q)]$ and the initial set of adopters grows linearly with the total population n.

Given a distribution $\boldsymbol{p} = (p_s)_{s\in\mathbb{N}}$, we define the functions:

$$\begin{aligned} h(z; \alpha, \boldsymbol{p}) &:= (1-\alpha)\sum_s p_s \sum_{r \geq s - \lfloor sq \rfloor} r b_{sr}(z), \\ g(z; \alpha, \boldsymbol{p}) &:= \lambda z^2 - h(z; \boldsymbol{\alpha}, \boldsymbol{p}), \\ h_1(z; \alpha, \boldsymbol{p}) &:= (1-\alpha)\sum_s p_s \sum_{r \geq s - \lfloor sq \rfloor} b_{sr}(z). \end{aligned}$$

We define

$$\hat{z}(\alpha, \boldsymbol{p}) := \max\{z \in [0,1] : \ g(z; \alpha, \boldsymbol{p}) = 0\}.$$

Proposition 3. *Consider the random graph $G(n, \mathbf{d})$ for a sequence $(d_i)_1^n$ satisfying Condition 1. If the size of the set of initial adopters is αn, then the final number of buyers is given by $(1 - h_1(\hat{z}, \boldsymbol{\alpha}, \boldsymbol{p}))n + o_p(n)$ provided $\hat{z}(\alpha, \boldsymbol{p}) = 0$, or $\hat{z}(\alpha, \boldsymbol{p}) \in (0, 1]$, and further $g(z; \alpha, \boldsymbol{p}) < 0$ for any z in some interval $(\hat{z} - \epsilon, \hat{z})$.*

We refer to [5] for a discussion of this result.

3. REFERENCES

[1] B. Bollobás. *Random graphs*, volume 73 of *Cambridge Studies in Advanced Mathematics*. Cambridge University Press, Cambridge, second edition, 2001.

[2] D. Easley and J. Kleinberg. *Networks, Crowds, and Markets: Reasoning About a Highly Connected World*. Cambridge University Press, 2010.

[3] A. Galeotti and S. Goyal. Influencing the influencers: a theory of strategic diffusion. *RAND Journal of Economics*, 40(3):509–532, 2009.

[4] M. O. Jackson and L. Yariv. Diffusion of behavior and equilibrium properties in network games. *The American Economic Review*, 97(2), 2007.

[5] M. Lelarge. Diffusion and cascading behavior in random networks. *CoRR*, abs/1012.2062, 2010.

[6] M. Molloy and B. Reed. A critical point for random graphs with a given degree sequence. *Random Structures Algorithms*, 6(2-3):161–179, 1995.

[7] S. Morris. Contagion. *Rev. Econom. Stud.*, 67(1):57–78, 2000.

[8] D. J. Watts. A simple model of global cascades on random networks. *Proc. Natl. Acad. Sci. USA*, 99(9):5766–5771 (electronic), 2002.