Minimal contagious sets in random regular graphs

Guilhem Semerjian

LPT-ENS

08.05.2014 / Warwick

Work (in progress) with Alberto Guggiola

Outline

Outline

Infection models

Modelization of epidemic propagation of :

- illnesses
- adoption of a new product (marketing)
- failure of banks

Formalization : $G = (V, E)$ a graph on N vertices

$$\partial i = \{j | (i, j) \in E\} \text{ the neighbors of } i$$

$\underline{\sigma}^{(t)} = (\sigma_1^{(t)}, \ldots, \sigma_N^{(t)})$ the state of the vertices at time t

Example : threshold model

$\sigma_i^{(t)} \in \{0,1\}$ 0=inactive=susceptible 1=active=infected

deterministic, parallel, discrete time dynamics :

$$\sigma_i^{(t)} = \begin{cases} 1 & \text{if } \sigma_i^{(t-1)} = 1 \\ 1 & \text{if } \sigma_i^{(t-1)} = 0 \text{ and } \sum_{j \in \partial i} \sigma_j^{(t-1)} \geq l_i \\ 0 & \text{otherwise} \end{cases}$$

- active vertices remain active for ever (monotonicity)
- inactive vertices become active if their number of active neighbors gets larger than some threshold (l_i for site i)

remark : related to bootstrap (q-core) percolation

[Chalupta, Leath, Reich 79]

Evolution from a random initial condition

First question ("direct problem", relatively easy on random graphs) :

assuming $\underline{\sigma} = \underline{\sigma}^{(0)}$ drawn from a product measure
(the initial state of each vertex is chosen at random in an i.i.d. way)

- what is the time evolution of $\underline{\sigma}^{(t)}$?
- for monotonous dynamics, what is the final state (large time limit)?
- is the final state completely active ?
  $\Leftrightarrow$ is the initial condition a "contagious set" ?

for instance SIR on random graphs via differential equations
[Janson, Luczak, Windridge 13]

Evolution from a random initial condition

Example : threshold model with $l_i = l$
on a (large) $k+1$ random regular graph (locally a regular tree)

[Balogh, Pittel 07]

Bernouilli random initial condition with probability θ of active sites

P_t : probability that a vertex is active at time t

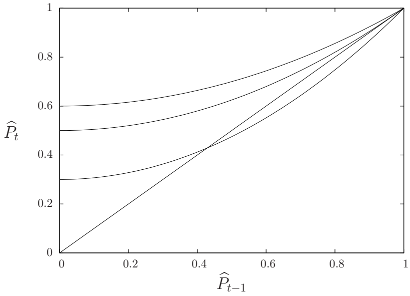
$$P_t = \theta + (1-\theta) \sum_{n=l}^{k+1} \binom{k+1}{n} (\widehat{P}_{t-1})^n (1-\widehat{P}_{t-1})^{k+1-n}$$

where $\widehat{P}_t$: probability that a vertex is active at time t, one of its neighbor being kept inactive

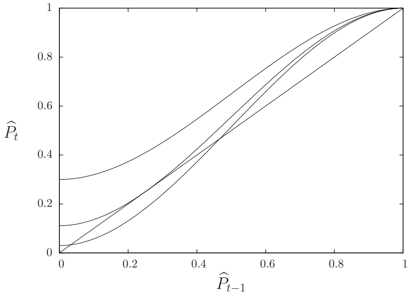
$$\widehat{P}_t = \theta + (1-\theta) \sum_{n=l}^{k} \binom{k}{n} (\widehat{P}_{t-1})^n (1-\widehat{P}_{t-1})^{k-n}$$

Evolution from a random initial condition

$l = k = 2$



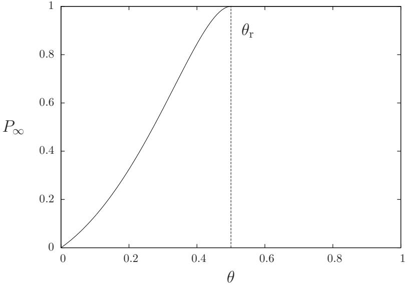
$l = 2, k = 3$



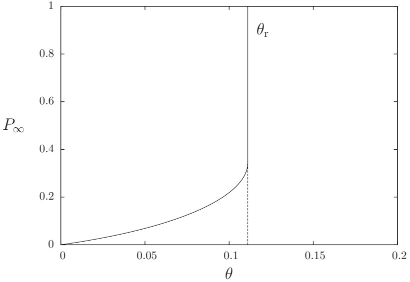
θ increases from bottom to top

Evolution from a random initial condition

$l = k = 2$



$l = 2, k = 3$



$\theta_\mathrm{r}(l,k)$: minimal value of θ for complete activation (w.h.p.) from a random Bernouilli condition

for $l = k$, $\theta_\mathrm{r} = \frac{k-1}{k}$, continuous transition

for $l < k$, discontinuous transition

Inference and optimization problems

Many other (more difficult) questions :

- inference

  given a snapshot $\underline{\sigma}^{(t)}$ of the epidemy at time t, infer what happened previously ($\underline{\sigma}^{(t')}$ with $t' < t$)

  in particular "zero patient" of an illness propagation

[Shah, Zaman 11]
[Pinto, Thiran, Vetterli 12]
[Lokhov, Mézard, Ohta, Zdeborova 13]
[Altarelli, Braunstein, Dall'Asta, Lage-Castellanos, Zecchina 13]

Inference and optimization problems

- optimization
  - which vertices should be vaccinated to minimize the propagation of an illness ? [Altarelli, Braunstein, Dall'Asta, Wakeling, Zecchina 13]
  - which vertices should be infected initially to maximize the propagation ? (viral marketing).
    For instance for the threshold model :
    - among the initial conditions with L active sites, find the one that maximizes the number of active sites in the final configuration "Maximum L influence" [Kempe, Kleinberg, Tardos 03]
    - find the initial configuration with the minimal number of active sites that activate the whole graph (the minimal contagious set) "Minimum Target Set Selection" [Chen 09]

Problems in general difficult (NP-hard) from the worst-case complexity (even to approximate)

Rigorous results on $\mathbb{Z}^d$, none (?) on random graphs apart from :

Bounds for good expanders

[Coja-Oghlan, Feige, Krivelevich, Reichman 13]

Outline

Definition

Reminder : threshold l on a $k+1$ regular graph G with N vertices

$$\sigma_i^{(t)} = \begin{cases} 1 & \text{if } \sigma_i^{(t-1)} = 1 \\ 1 & \text{if } \sigma_i^{(t-1)} = 0 \text{ and } \sum_{j \in \partial i} \sigma_j^{(t-1)} \geq l \\ 0 & \text{otherwise} \end{cases}$$

$\sigma_i^{\mathrm{f}} = \lim \sigma_i^{(t)}$ as $t \to \infty$ (exists by monotonicity)

$\theta_{\min}(G) = \frac{1}{N} \min_{\underline{\sigma}} \{ \sum_i \sigma_i^{(0)} \mid \sigma_i^{\mathrm{f}} = 1 \ \forall i \}$

minimal fraction of active sites in an initial configuration that activates all vertices (i.e. minimal contagious set)

$\theta_{\min}(l,k) = \lim_{N \to \infty} \mathbb{E} \, \theta_{\min}(G)$ average over random regular graphs

limit should exist, and $\theta_{\min}(G)$ should concentrate around it

Definition

Variant of the problem : "time horizon" T

$$\theta_{\rm min}(G,T) = \frac{1}{N} \min_{\underline{\sigma}} \{ \sum_i \sigma_i^{(0)} \mid \sigma_i^{(T)} = 1 \ \forall i \}$$

minimal fraction of active sites in an initial configuration that activates all vertices in T steps (minimal contagious set corresponds to $T = \infty$)

$$\theta_{\rm min}(l,k,T) = \lim_{N \to \infty} \mathbb{E}\, \theta_{\rm min}(G,T)$$ average over random regular graphs

here limit and concentration can be proven by interpolation method

$$\theta_{\rm min}(l,k) \le \lim_{T \to \infty} \theta_{\rm min}(l,k,T)$$ (could be equal)

Different views on the problem

- packing problem : put as many inactive sites as possible in the initial configuration, under some constraints :
  - for $l = k + 1$ (and any T) : maximal independent set (hardcore model)
    two neighboring inactive vertices will never evolve
  - for $T = 1$, Biroli-Mézard model : one inactive site can have at most $k + 1 - l$ inactive neighbors

Different views on the problem

- Avoided (core) percolation

  activate a site with $\geq l$ active neighbors $\Leftrightarrow$
  remove an inactive site with $< k - l + 2$ inactive neighbors

  in $\underline{\sigma}^{\rm f}$, the inactive sites form the $(k - l + 2)-$core (largest subset with induced degree $\geq k - l + 2$) of the inactive sites in the initial configuration

  $1 - \theta_{\rm min}$: maximal density of vertices avoiding core percolation

  reminiscent of offline version of Achlioptas processes
  [Bohman, Frieze, Wormald 04]
  cf. Lutz Warnke talk on monday

Statistical mechanics formulation

Statistical mechanics formulation (for finite T) :

$$\eta(\underline{\sigma}) = \frac{1}{Z} e^{\mu \sum_i \sigma_i} \prod_i \delta_{\sigma_i^{(T)},1} \ , \quad Z = \sum_{\underline{\sigma}} e^{\mu \sum_i \sigma_i} \prod_i \delta_{\sigma_i^{(T)},1}$$

μ : “chemical potential”, “activity”, minimal sets when $\mu \to -\infty$:
the partition function Z contains the relevant information :

$$\lim_{\mu \to -\infty} \frac{1}{N\mu} \ln Z = \theta_{\min}(G,T) \text{ and even more precisely :}$$

- define the entropy $s_T(\theta)$ as $(1/N) \ln(|$ activating initial configurations with θN active vertices$|)$
- then $\frac{1}{N} \ln Z \approx \sup_\theta [s_T(\theta) + \mu\theta]$ (Legendre transform)

can one compute Z ?

Statistical mechanics formulation

$$\eta(\underline{\sigma}) = \frac{1}{Z} e^{\mu \sum_i \sigma_i} \prod_i \delta_{\sigma_i^{(T)},1} \ , \quad Z = \sum_{\underline{\sigma}} e^{\mu \sum_i \sigma_i} \prod_i \delta_{\sigma_i^{(T)},1}$$

$\sigma_i^{(T)}$ function of all σ_j at distance smaller than T, non local interactions, not convenient

$\Rightarrow$ reformulation with additional (redundant) time variables

[Altarelli, Braunstein, Dall'Asta, Zecchina 12]

$t_i(\underline{\sigma})$: first time at which i is active starting from $\underline{\sigma}$ (or ∞ if $> T$)

$$t_i(\underline{\sigma}) = \begin{cases} 0 & \text{if } \sigma_i = 1 \\ 1 + \min_l(\{t_j(\underline{\sigma})\}_{j \in \partial i}) & \text{if } \sigma_i = 0 \text{ and } 1 + \dots \leq T \\ \infty & \text{otherwise} \end{cases}$$

where $\min_l(t_1, \dots, t_n) = t_l$ if $t_1 \leq \dots \leq t_n$

Statistical mechanics formulation

One can thus rewrite

$$Z = \sum_{\underline{t}} e^{\mu \sum_i \delta_{t_i,0}} \prod_i \mathbb{I}(t_i \leq T) \prod_i w_i(t_i, \{t_j\}_{j \in \partial i})$$

where the w_i encode the local consistency between activation times

Turned into a model with variables $t_i \in \{0, \dots, T, \infty\}$ on each vertex, with local interactions

Cavity method has been precisely developed to compute Z for such models, when the graph is a sparse random graph

(local convergence towards a tree)

[Mézard, Parisi 00]

Outline

The cavity method

Simplest example : Ising ferromagnet on a $k+1$ random regular graph

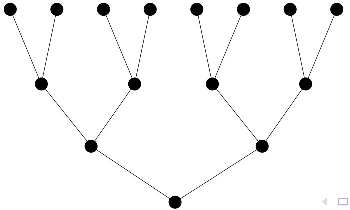
Ising spins on the vertices : $\underline{\sigma} = (\sigma_1, \dots, \sigma_N) \in \{-1,+1\}^N$

interaction along the edges : $H(\underline{\sigma}) = -J \sum_{\langle i,j \rangle \in E} \sigma_i \sigma_j$

$$\eta(\underline{\sigma}) = \frac{1}{Z} e^{-\beta H(\underline{\sigma})} \ , \qquad Z = \sum_{\underline{\sigma}} e^{-\beta H(\underline{\sigma})}$$

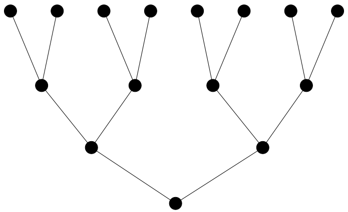
random graphs converge locally to trees

models on finite trees are simple



The cavity method

$$H(\sigma_1, \ldots, \sigma_N) = -J \sum_{\langle i,j \rangle} \sigma_i \sigma_j \ , \qquad Z = \sum_{\sigma_1=\pm 1} \cdots \sum_{\sigma_N=\pm 1} e^{-\beta H(\sigma_1,\ldots,\sigma_N)}$$

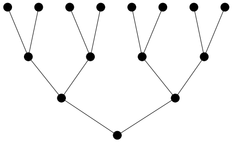


$Z_g(\sigma)$: partition function

- conditioned on the value of the root σ
- in a regular tree with g generations

The cavity method

$$Z_{g+1}(\sigma) = \sum_{\sigma_1,\dots,\sigma_k} Z_g(\sigma_1)\dots Z_g(\sigma_k)\, e^{\beta J \sigma(\sigma_1+\dots+\sigma_k)}$$



Normalized probability : $\eta_g(\sigma) = \frac{Z_g(\sigma)}{Z_g(+)+Z_g(-)} = \frac{e^{\beta h_g \sigma}}{2\cosh(\beta h_g)}$

Recursion on the effective magnetic field :

$$h_{g+1} = \frac{k}{\beta} \text{atanh}\left(\tanh(\beta J)\tanh(\beta h_g)\right)$$

Fixed point when $g \to \infty$: (infinitesimal field to break the symmetry)

- $h = 0$ at high temperature
- $h \neq 0$ at low temperature

True magnetic field can be recovered
with $k+1$ instead of k neighbors on the root

The cavity method

- Generalization to any model on a finite tree :

  solvable via exchange of “messages” between neighboring variables (Belief Propagation)

- Generalization to models on random graphs :

  only locally tree-like, effect of the loops (boundary conditions)

  - in simple cases, Replica Symmetric phase, fast correlation decay
    - Ferromagnetic Ising models [Dembo, Montanari 08]
    - Matchings [Bordenave, Lelarge, Salez 11]
  - in more complicated cases, Replica Symmetry Breaking, correlated boundary conditions, computation of the number of clusters (pure states)

Method applicable to any model with an interaction graph converging locally to a tree, random constraint satisfaction problems, lattice glasses, properties of random graphs...

The cavity method

lots of (heuristic) predictions from statistical physics
for various random graph problems

in particular random constraint satisfaction problems (k-SAT, q-COL)
[Mézard, Parisi, Zecchina 03]
[Krzakala, Montanari, Ricci-Tersenghi, Semerjian, Zdeborova 07]

Two recent rigorous confirmations :

- largest independent sets of random regular graphs
[Ding, Sly, Sun 13]
- k-SAT threshold [Coja-Oghlan 13]

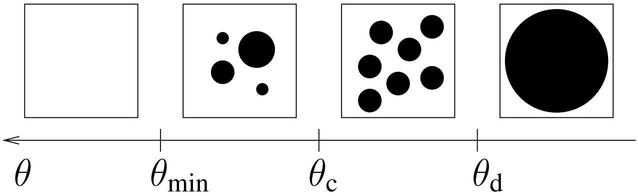
Outline

Results

Back to the contagion problem :

$$\eta(\underline{\sigma}) = \frac{1}{Z} e^{\mu \sum_i \sigma_i} \prod_i \delta_{\sigma_i^{(T)},1} \ , \quad Z = \sum_{\underline{\sigma}} e^{\mu \sum_i \sigma_i} \prod_i \delta_{\sigma_i^{(T)},1}$$

For all k, l and finite T, structure of the support of η :



One step of replica symmetry breaking (1RSB) scenario

Results

To compute numerically the 1RSB prediction for $\theta_{\min}(k, l, T)$:

- there are two functions $\Sigma(y)$ and $\theta(y)$, that can be computed from the solution of a set of $2T$ algebraic equations on $2T$ unknowns, dependent on the parameter $y \in [0, \infty[$
- find y_* such that $\Sigma(y_*) = 0$
- return $\theta(y_*)$

Lots of numbers as a function of k, l, T...

The limit $T \to \infty$ can be performed analytically, in the sense that $\Sigma(y)$ and $\theta(y)$ can be computed from the solution of a finite set of algebraic equations

Yields qualitatively different results depending on k and l

Results

Simplest case : $k = l = 2$
(i.e. 3-regular random graph with activation threshold 2)

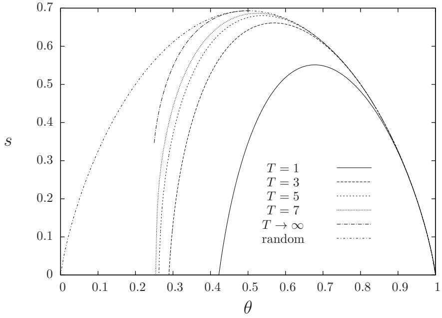
Conjecture :

$$\begin{aligned}\lim_{T\to\infty} \theta_{\min}(k = l = 2, T) &= \lim_{T\to\infty} \theta_{\mathrm{c}}(k = l = 2, T) \\ &= \lim_{T\to\infty} \theta_{\mathrm{d}}(k = l = 2, T) \\ &= \frac{1}{4} = \frac{1}{2}\,\theta_{\mathrm{r}}(k = l = 2)\end{aligned}$$

such minimal contagious sets can be easily found by a greedy algorithm

Results

Further conjecture on the entropy in the $T \to \infty$ limit for $l = k = 2$:



- for $\theta > \theta_{\rm r}$, one has $s(\theta) = -\theta \ln \theta - (1-\theta)\ln(1-\theta)$:

  typical random configurations do activate (trivial, true for all k, l)

- for $\theta \in [\theta_{\rm r}/2, \theta_{\rm r}]$,
  $s(\theta) = -\left(2\theta - \frac{1}{2}\right)\ln(2\theta - \frac{1}{2}) + 2\theta \ln \theta + \frac{3}{2}\ln 2 \qquad s(\theta_{\min}) = (\ln 2)/2$

Results

Second simplest case : $k = l = 3$

$\theta_{\rm min}(k = l = 3, T) \underset{T\to\infty}{\to} \frac{1}{3} = \frac{1}{2}\,\theta_{\rm r}(k = l = 3)$

but such minimal contagious set will not be found by a simple greedy algorithm

for $l = k \geq 4$ and (most) $l < k$ cases, no simple prediction for $\theta_{\rm min}$

Analytical expansion at large k to be performed

Conclusions

- one among many other examples where random graph problems can be treated with statistical mechanics methods (matchings, independent sets, random constraint satisfaction problems...)
- conjectures awaiting for a rigorous (dis)proof
- if/when true (stability of the 1RSB scenario), same flavour as the recent rigorous results on largest independent sets of random regular graphs [Ding, Sly, Sun 13]
  and k-SAT threshold [Coja-Oghlan 13]