

Detecting modules in co-expression networks across species

KKY, Nets Apr, 2012



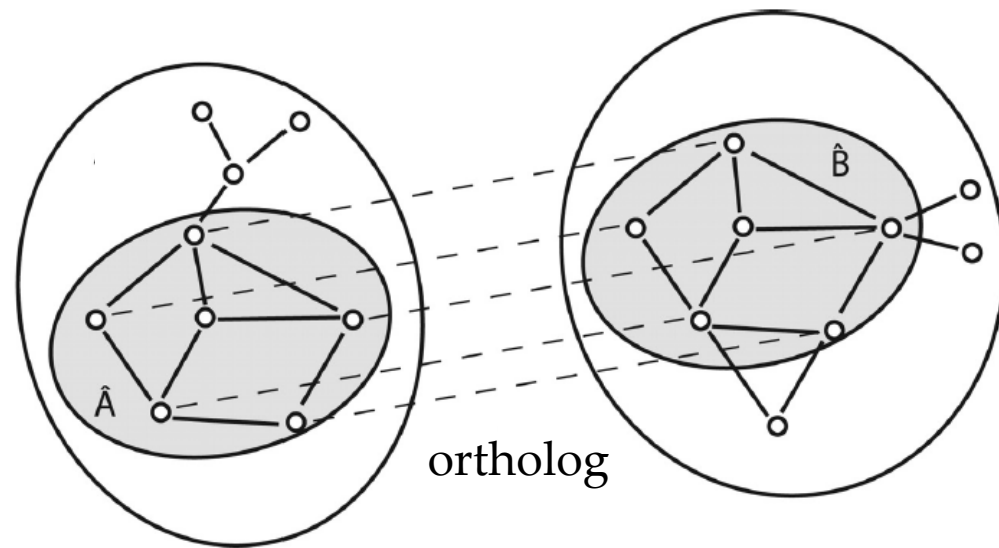
Motivation

- There are well known packages (WGCNA) to detect co-expressed modules in a co-expression, but evolutionary information has not been incorporated
- Given two co-expressed networks. If a pair of orthologs perform the same function in two species, their corresponding modules should participate the same biological processes. Evolutionary information could therefore be used to refine co-expressed modules, in addition to the correlation between genes within a network.



Motivation

- In this language, given two co-expressed networks A1 and A2, a module is a set of nodes, M1 of them from A1 and M2 from A2. The expression profiles of M1 (M2) genes are highly correlated, and there are many orthologs between M1 and M2



Potts Model

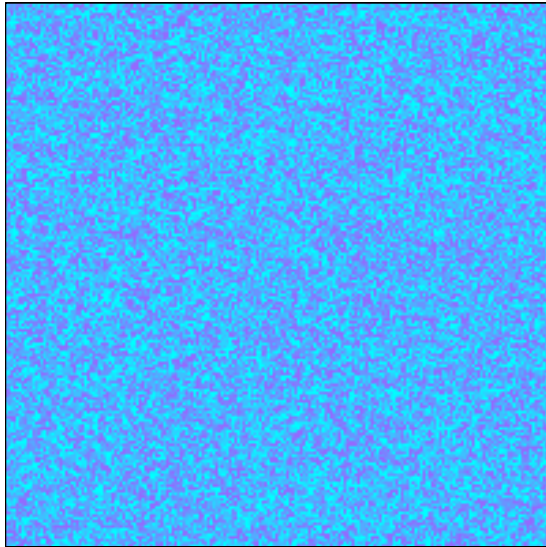
- In a lattice, each site takes on values ranging from $1, 2, \dots, q$. The energy of then system is

$$H = - \sum_{(i,j) \in nn} \delta_{\sigma_i \sigma_j} + \gamma \sum_{m=1}^q \frac{n_m (n_m - 1)}{2}$$

among the nearest neighbors j ,
how many of them have the
same spin as site i

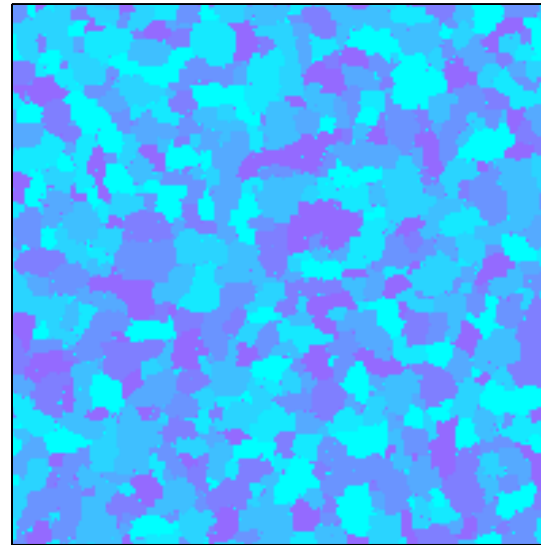
anti-ferromagnetic term that
avoid too many sites from
having the same spin value

200x200 Potts model with 8 spins



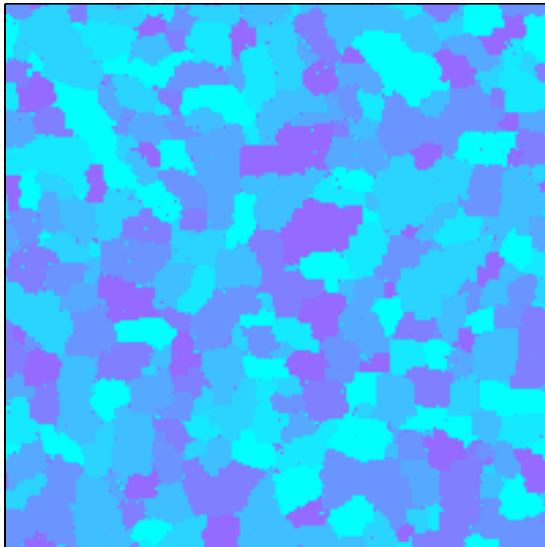
Step = 1
 $t=0$

200x200 Potts model with 8 spins



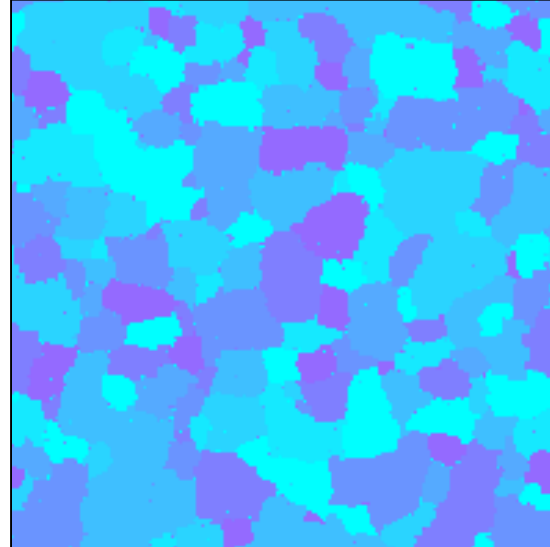
Step = 100
 $t=100$

200x200 Potts model with 8 spins



Step = 200
 $t=400$

200x200 Potts model with 8 spins



Step = 400
 $T=800$

Using Potts model to detect modules in a network

- Cliques in a network are likely to end up with the same value of spin, and thus classified into a module.

Protein complexes and functional modules in molecular networks

Victor Spirin and Leonid A. Mirny*

PNAS 2003

VOLUME 93, NUMBER 21

PHYSICAL REVIEW LETTERS

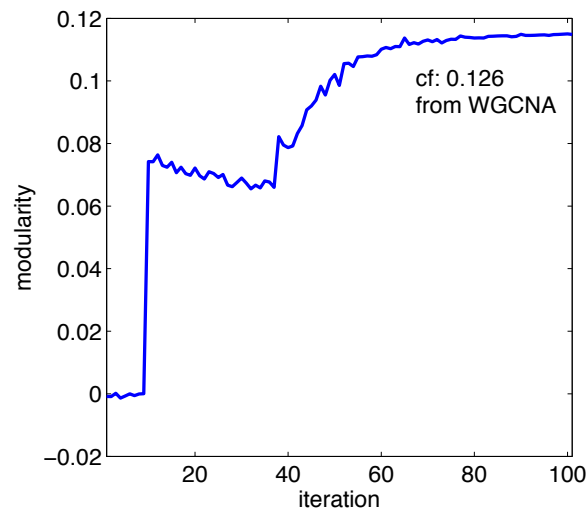
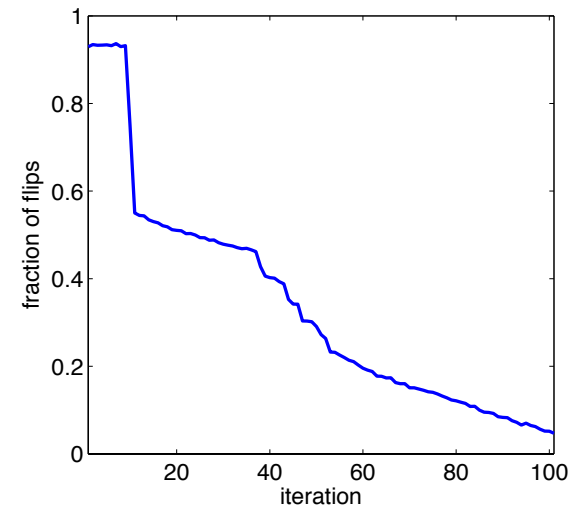
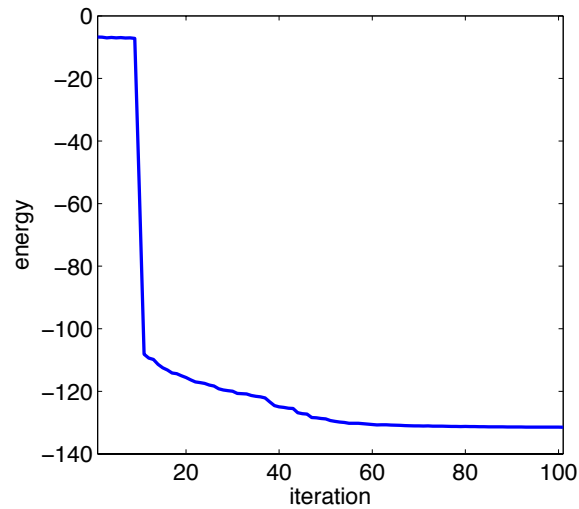
week ending
19 NOVEMBER 2004

Detecting Fuzzy Community Structures in Complex Networks with a Potts Model

Jörg Reichardt* and Stefan Bornholdt*

Test drive

co-expressed network:
960 nodes, 90000 edges



beta start=0.025;
gamma=0.2
In each iteration
sweep through the network 20 times
then beta increases by 5%

74% agreement

Coupled Potts model

Given two networks, size N_1 and N_2

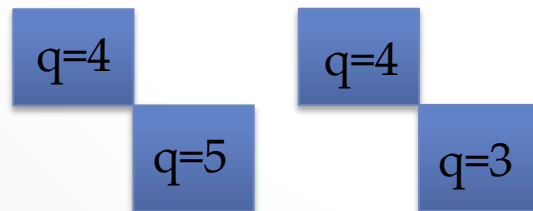
$$H = \left(- \sum_{(i,j) \in E_1} \delta_{\sigma_i \sigma_j} + \gamma \sum_{m=1}^q \frac{n_m^{(1)}(n_m^{(1)} - 1)}{2} \right) \frac{1}{N_1} + \left(- \sum_{(i',j') \in E_2} \delta_{\sigma_{i'} \sigma_{j'}} + \gamma \sum_{m=1}^q \frac{n_m^{(2)}(n_m^{(2)} - 1)}{2} \right) \frac{1}{N_2}$$

$$- \sum_{(i,j') \in Ortho} \delta_{\sigma_i \sigma_{j'}} \frac{N_1 + N_2}{2N_1N_2} k$$

Two coupling constants
to be tuned: γ, k

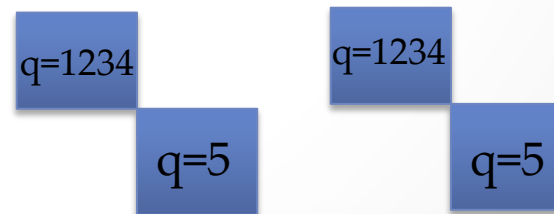
Test Drive

- Construct a network A1 (N=200), with 2 modules (100 nodes each), within each modules, connections are denser than average
- Make another network A2, by rewiring 5% of A1 edges.
- Assume nodes in A1 and A2 form orthologs
- No coupling. Beta=2, gamma=1, iteration=50, q=5
 - A1: For nodes in module 1, 99 of them have value 4; for nodes in module 2, 99 of them have value 1
 - A2: For nodes in module 1, 99 of them have value 1; for nodes in module 2, 97 of them have value 5
- With coupling. K=1



101/200 agrees

K=10



199/200 agrees

How to justify the output?

- This is the difficult part. Also closely related to the tuning of coupling parameters.
- The output of the algorithm (a module), is two sets of nodes from two species, which presumably corresponding to the same biological function.
 - GO enrichments for two sets respectively, check similarities

Questions to address

- Given a module. two sets of nodes from two species
 - How often do the two sets share common orthologs?
 - How often do they acquire species specific genes?
- In principles, not restricted to co-expressed networks, but say PPI networks