Published in final edited form as:

IEEE/ACM Trans Comput Biol Bioinform. 2015 ; 12(3): 622–631. doi:10.1109/TCBB.2014.2366748.

Evolutionary model selection and parameter estimation for protein-protein interaction network based on differential evolution algorithm

Lei Huang,

Department of Computer and Information Sciences, University of Delaware, Newark, DE, 19716. leihuang@udel.edu

Li Liao, and

Department of Computer and Information Sciences, University of Delaware, Newark, DE, 19716. liliao@udel.edu

Cathy H. Wu

Delaware Biotechnology Institute, University of Delaware, Newark, DE, 19711. wuc@dbi.udel.edu

Abstract

Revealing the underlying evolutionary mechanism plays an important role in understanding protein interaction networks in the cell. While many evolutionary models have been proposed, the problem about applying these models to real network data, especially for differentiating which model can better describe evolutionary process for the observed network urgently remains as a challenge. The traditional way is to use a model with presumed parameters to generate a network, and then evaluate the fitness by summary statistics, which however cannot capture the complete network structures information and estimate parameter distribution.

In this work we developed a novel method based on Approximate Bayesian Computation and modified Differential Evolution (ABC-DEP) that is capable of conducting model selection and parameter estimation simultaneously and detecting the underlying evolutionary mechanisms more accurately. We tested our method for its power in differentiating models and estimating parameters on the simulated data and found significant improvement in performance benchmark, as compared with a previous method. We further applied our method to real data of protein interaction networks in human and yeast. Our results show Duplication Attachment model as the predominant evolutionary mechanism for human PPI networks and Scale-Free model as the predominant mechanism for yeast PPI networks.

Index Terms

Minimax approximation and algorithms; Eigenvalues and eigenvectors; Probabilistic algorithms; Convergence

1 Introduction

Protein-protein interactions (PPIs), imply physical contacts between two or more proteins as a result of biochemical events and/or electrostatic forces, which is vital to understanding

protein function within the cell. More importantly, almost all the biological processes are regulated through interactions or isolations between protein molecules. In recent years, an increasing number of PPIs have been detected due to the advances of experimental methods and bioinformatics methods. Consequently, more and more researchers begin doing research at PPIs' network level, more specifically, focusing on the evolutionary mechanism of PPI network. Several popular evolutionary models used to simulate the evolution process of PPI network have been proposed in the past few years. It is commonly believe that one mechanism by which PPIs network evolve is gene duplication, subsequently, another mechanism named post-duplication divergence may cause the PPI network further evolve. Briefly, the whole process can be described as given an original PPIs network, a node may be duplicated at a certain probability; and then during the divergence process, some new connections may be formed between the duplicated node and existing nodes, meanwhile, some existing edges may be deleted. Based on the duplication and divergence mechanisms, many evolutionary models have been proposed[1], [2], [3] among which post-duplication divergence is the main difference. Besides, some early studies suggested scale free[4] model may fit PPI network well[5], [6], but there are several statistical challenges for this claim[7], [8].

With the increasing number of evolutionary models, it is urgent to develop accurate analysis methods for evaluating the fitness of evolutionary models. Traditionally, researchers would like to evaluate the difference between simulation network and observed network on the basis of topological features, such as degree[9], [10], betweenness[11], modularity[12], diameter[13], clustering coefficient[14], assortativity[15], [16] and so on. But it is difficult to describe the PPIs network in terms of these summary statistics for the noise and incompleteness. To deal with this problem, most recently, Thorne, T. and Stumpf, M. P.[17] developed Approximate Bayesian computation and sequential Monte Carlo method (ABC-SMC) to do graph spectral analysis, which enables model selection and parameter estimation over a number of network evolutionary models. It has been demonstrated that the graph spectra based ABC-SMC can capture network data more naturally than the traditional summary statistics. However, it cannot differentiate similar models accurately, especially for many duplication-divergence based models. Meanwhile, it does not quantitatively analyze the difference between simulation network and observed network. Posterior probability alone is not convincing enough to do model selection. Moreover, for each time, the sequential Monte Carlo sampling based ABC-SMC needs to choose a proper threshold value ε that is used to accept or reject a particle. It is however hard to select the right value, if ε is too large, it may take too long to find the good particles. If the ε is too small, it will result in many particles being drawn that are never used. So the efficiency of ABC-SMC method is largely restricted by the choice of ε. After reproducing the method, we found the ABC-SMC method indeed converges slowly and fluctuantly.

To deal with these issues, we propose an improved graph spectra analysis method based on approximate Bayesian computation with differential evolution method (DE)[18] and propagation (ABC-DEP). DE is demonstrated one of best methods for optimization problems. Moreover, to make DE more suitable for the evaluation of posterior density over a number of models, we combine it with an additional propagation kernel. The experimental

results show our method can differentiate similar evolutionary models accurately. And some quantitative analysis demonstrates our method converge rapidly and smoothly.

In the method section, we give a detail introduction about our method. We demonstrate the accurateness, robustness, and reliability by testing ABC-DEP based on simulation networks. To show the promising ability of ABC-DEP, we apply it to PPIs network downloaded from PrePPI database[19], [20]. Finally, we conclude by discussing results and emphasizing the significance of our method.

2 Methods

In this section, we introduce several key parts of our method first, and then outline ABC-DEP framework.

2.1 Approximate Bayesian computation (ABC)

Given an observed PPI network, D, and a set of evolutionary models m_i with parameters θ, we develop an efficient method that can carry out model selection and parameter estimation simultaneously to detect the underlying evolutionary mechanism. Being a probabilistic approach, our method is based on the Bayesian analysis to compute the posterior probability of any model m_i, given a network D :

$$p(m_i(\theta)|D)=\frac{p(D|m_i(\theta))p(m_i(\theta))}{p(D)} \quad (1)$$

Where $p(D|m_i(\theta))$ the likelihood, $p(m_i(\theta))$ the prior, and $p(D)$ the evidence. The prior $p(m_i(\theta))$ is assumed to be known and often is specified by choosing a particular distribution; here uniform distribution has been chosen for our method, such that both evaluation of prior probabilities and random generation of value θ are relatively straightforward. However, it is computationally expensive, or even totally impossible to evaluate the likelihood $p(D|m_i(\theta))$.

Therefore, we choose to not evaluate the likelihood precisely, but do an approximate Bayesian computation (ABC) instead. All ABC based methods approximate the likelihood by simulations whose outputs (simulation network) are compared with the observed network[21]. More specifically, a set of parameters for a certain model is sampled through a presumed prior distribution. The model and its parameters form a so-called ”particle” $m_i(\theta)$ with which we can simulate a network D'. Then we would like to evaluate the distance between simulation network D' and the observed network D in order to accept or discard this particle. If the distance is smaller than a preset threshold, the sampled particle will be accepted, otherwise, will be discarded. The basic formula can be given by:

$$p(m_i(\theta)|D) \approx p(m_i(\theta)|d(D', D)<\varepsilon) \quad (2)$$

Where ε represents the threshold to judge the distance, and $d(D', D)$ represents the distance between network D' simulated by particle $m_i(\theta)$ and observed network D.

2.2 Distance computing method for networks

As described in last subsection, we need to evaluate the distance between the simulated network and the observed network. To begin with, we represent a network by adjacency matrix that is supposed to capture all the structure information of the network, if only implicitly. Given a network with *N* nodes and *E* edges, the corresponding adjacency matrix *M* with $N \times N$ dimension can be given by:

$$a_{i,j} \begin{cases} 1, if(i,j) \in E \\ 0, if(i,j) \notin E \end{cases} \quad (3)$$

Where *i* and *j* are two nodes in the nodes set *N*, and (*i*, *j*) represents an edge between *i* and *j*, $(i, j) \in E$. Suppose the simulation network D' and the observed network *D* are represented by matrices *A* and *B* respectively. In theory, we may just compute the distance between *A* and *B* by Eq.(4), where $a_{i,j}$ and $b_{i,j}$ are elements in matrix *A* and *B*.

$$d(A,B) = \sum_i \sum_j (a_{i,j} - b_{i,j})^2 \quad (4)$$

However, the PPI networks are usually unlabeled and undirected, which means we do not know the mapping strategy between the simulation network and observed PPI network. Considering the PPI networks are with large size while very sparse, for example, the human high-confidence PPI network[19], [20] has 4003 nodes and 6780 edges, so it is extremely expensive to get either the minimum or the average distance between the observed PPI network and the corresponding simulation network by permuting all possible mapping strategies between them.

Therefore, instead of the naive method, we adopt the theorem of Umeyama[22] by which the approximate lower bound on the edit distance between two networks can be obtained. It has been demonstrated in Wilson and Zhu[23] that the lower bound on edit distance is an excellent approximation of edit distance between two networks. The measure formula is shown by

$$d(A,B) = \sum_i (\alpha_i - \beta_i)^2 \quad (5)$$

Where *A* and *B* are Hermitian matrices, and α_i and β_i are their ordered eigenvalues respectively. We will further demonstrate the reliability of Eq.(5) when we do distance analysis in the later subsection.

2.3 Differential Evolution algorithm

Differential evolution (DE) is a population based, stochastic function minimizer, which is shown to be the best genetic type of algorithm for solving the real-valued test function suite of the First International Contest on Evolutionary Computation[24]. It has been widely applied to optimization problems of different kinds in various research fields. DE has been

adopted as the foundation of our ABC-DEP algorithm for its efficiency, accuracy and reliability.

Briefly, the central idea behind DE is a self-organizing scheme for generating trial parameter vectors by mutation and crossover, and then the trial vector will be selected or discarded by an objective function. Fig.1 shows the more detailed process of DE algorithm. Given a population of particles, a target vector, a randomly chosen base vector and another two different random vectors are needed to do mutation that is adding the weighted difference between the two random vectors to the base vector. After that, a crossover between the mutant vector and the target vector is used to generate a trial vector. Finally, a choice between target vector and trial vector is made by evaluating their objective function value. Typically, the whole process needs to be repeated multiple times in order to get the optimization output.

2.4 ABC-DEP for model selection and parameter estimation

Before introducing the ABC-DEP algorithm, what should be mentioned in advance is that we treat models and parameters analogously and encode the different models as another parameter in order to do model selection and parameter estimation simultaneously in one evolution procedure, which is inspired by the method of Toni and Stumpf[25] and Thorne and Stumpf[17]. As mentioned in the previous section, DE is an excellent method for solving the optimization problem. The problem, however, we need to solve is to do model selection and parameter estimation by evaluating the posterior probability, which is based on importance sampling. We make two-tuples particle that consist of a certain model and its parameter vector as a member of population. The DE algorithm may help us find several good particles, but what we need is the posterior distribution of particles. To address this issue, as illustrated in Fig.2, we propose another evolution kernel, propagation, and combine it with DE.

2.4.1 Initialization—To do initialization, we randomly choose one out of the six evolutionary models and then randomly assign values from a preset range to the parameters for this model, and make the model and its parameters into a particle by which we can generate a simulated network D'. Next, we evaluate the distance between the simulation network and the observed network D. If $d(D', D) < \varepsilon$, we accept this particle and assign it an initial weight $w = 1$, otherwise, it is discarded. Here ε is a threshold used to control the quality of particles in a fixed scale of population. Different from ABC-SMC[19] that highly depends on a set of thresholds ε to guarantee the accuracy of final posterior probability, we only need one threshold ε for the initialization. The initiation process will be carried out repeatedly until N particles have been accepted into the population.

2.4.2 Evolution—Once N particles have been initialized, the procedure goes into the evolution part that includes two kernels: DE and Propagation. As we mentioned in previous sections, we need to randomly select other three different particles for the target particle to do mutation. However, the DE kernel tends to reduce the diversity of the population after several iterations, making it difficult to find other three different particles with the same model but different parameters, which are used to form a mutant to crossover with the target

particle. When this happens, ABC-DEP will switch into the propagation kernel. Propagation means, for the sole target particle without finding three other different particles to form a mutant, we will use a multivariate Gaussian distribution with zero mean and a proper variance to perturb its parameter vector. As illustrated in Fig.3, while DE can help find several optimized particles, propagation enables us to find some good neighbors around the optimized particles and diversify the population.

Each particle in the population will be selected as a target particle to do evolution. Then, we can obtain a trial population that has the same size of the target population. For every pair of particles, one from each of these two populations, we will use them to generate a simulate network respectively. Suppose network D'_{P^i} is simulated by a target particle P^i, and D'_{Z^i} is simulated by the trial particle Z^i, where Z^i is evolved from P^i. Then ABC-DEP adopts Metropolis-hasting acceptance probability to determine whether the trial particle Z^i should be accepted to replace the target particle P^i. The Metropolis-hasting acceptance probability can be evaluated by (6)

$$min(1, \frac{d(D, D'_{P^i}) t(P^i|Z^i)}{d(D, D'_{Z^i}) t(Z^i|P^i)}) \quad (6)$$

Where the distance $d(D', D)$ is used to evaluate the fitness of simulation network, and $t(b|a)$ represents the transfer probability from a to b. Here we simplify the problem by not differentiating the propagation direction. That is, $t(Z^i|P^i)$ and $t(P^i|Z^i)$ are equal; they are given by the following formula.

$$t(Z^i|P^i) = t(P^i|Z^i) = CR \sum_i \sum_j \sum_k P^i.\theta + F(P^j.\theta - P^k.\theta) \quad (7)$$

CR is the crossover probability and *F* is the weighting factor in DE. On the basis of (6) and (7), therefore, the Metropolis-hasting acceptance probability is

$$min(1, \frac{d(D, D'_{P^i})}{d(D, D'_{Z^i})}) \quad (8)$$

The trial particle may be accepted with the probability given in Eq.(8) to replace the target particle; otherwise, the target particle will be kept for the next generation. Next, ABC-DEP updates particle's weight by the method shown in Algorithm DEP. The importance of updating weight by multiplying $\frac{\varepsilon}{D', D}$ is to incorporate the fitness of the particle, namely, the better (i.e., with a smaller *d*) the simulated network is, the higher the weight of that particle is. Unlike the ABC-SMC[17] which may inefficiently try hundreds or thousands times to get a satisfying particle for the continuously strict acceptance threshold ε and which is a problem that becomes especially serious during the last few iterations, our method only need one time to select a particle based on Eq.(8) for the next generation.

2.4.3 Sampling—Before sampling, the weights of the N_p particles through evolution are normalized first. Then a model's intermediate posterior probability can be obtained by adding the weights of its particles. For instance, there are n particles that belongs to model i, $n < N_p$ then the model i's posterior probability is given by

$$Pro_{model[i]} = \sum_{j=1}^{n} w^j \quad (9)$$

During sampling, a model is selected based on its intermediate posterior probability, and then for the selected model a specific particle is chosen based on the particle's weight, i.e., the particle with higher weight and belongs to a model with higher posterior probability will get more chance to survive for the next generation. Besides, in order to prevent a certain model from being extinct, for the model selection, it may stay at the selected model at probability p, or randomly jump to other model at probability $1 - p$. Therefore, we assign the new sampled particle with a weight by

$$w_{particles \in model[i]} = p \times Pro_{model[i]} + (1-p) \sum_{j \neq i} Pro_{model[j]} \quad (10)$$

Where $(1 - p)$ represents the transfer probability. This sampling procedure should be continued until N_p particles have been sampled. We do evolution and sampling repeatedly without decreasing the acceptance threshold ε until converged model posterior probabilities are obtained.

Algorithm 1

ABC-DEP

Require: $Model_i \leftarrow$ *evolution models*
$M \leftarrow$ *iterations times*
$N \leftarrow$ *particles*

while $t \leq M$ **do**
if $t = 1$ **then**
Initialize N particles satisfy $d(D, D') < \varepsilon$
$P_{particles} \leftarrow \{P^i, W^i\}_{i=1}^{N}$
else
$\{P_t, W_t\}_{i=1}^{N} \leftarrow Sampling((P_{t-1}, W_{t-1}))$
end if
$(P_{t+1}, W_{t+1}) \leftarrow DEP(P_t, W_t)$
$t \leftarrow t + 1$
end while
$Normalize((P_M, W_M))$
$PosteriorPro(Model_i, (P_M, W_M))$

3 Results and discussion

Before applying ABC-DEP to real PPI networks, we tested our method first on simulated data. To compare with the work reported in T. Thorne and M. P. H. Stumpf[17], the same six evolution models, Duplication Attachment (DA), Duplication Attachment with Complementarity (DAC), Linear Preferential Attachment (LPA)[4], the general Scale Free (SF)[26], Combination model of DAC and LPA (DACL) and DAC model with random edges addition(DACR), are included in our experiment. The experiment based on simulated data aims to evaluate how accurately ABC-DEP can predict the underlying model that is used to simulate the test data. Finally, we use ABC-DEP to analyze the possible evolutionary mechanism for real PPI networks.

Algorithm 2

DEP

Require: Population with size N_p

for $i = 1$ to N_p **do**

Randomly select P^f, P^j, P^k where $i \neq j \neq k \neq f$ and $P^i.m = P^j.m = P^k.m = P^f.m$

if $P^f.\theta = P^j.\theta = P^k.m = P^k.\theta$ **then**

$Z^i = Propagation(P^i)$

else

$Z^i = DifferentialEvolution(P^i, P^j, P^k, P^f)$

end if

end for

for $i = 1$ to N_p **do**

Simulate D'_{Z^i} by particle Z^i and D'_{P^i} by particle P^i

if $rand(0, 1) \leq min(1, \frac{d(D, D'_{P^i})}{d(D, D'_{Z^i})})$ **then**

if $\frac{d(D, D'_{P^i})}{d(D, D'_{Z^i})} \geq 1$ **then**

$$W_{P^i} \leftarrow W_{P^i} \times \frac{\varepsilon}{d(D, D'_{Z^i})}$$

else

$$W_{P^i} \leftarrow W_{P^i} \times \frac{d(D, D'_{P^i})}{d(D, D'_{Z^i})} \times \frac{\varepsilon}{d(D, D'_{Z^i})}$$

end if

$P^i \leftarrow Z^i$

else

$$P^i \leftarrow P^i \qquad W^{P^i} \leftarrow W^{P^i} \times \frac{\epsilon}{d(D, D'_{P^i})}$$

end if

end for

$Normalize(P^i, W^{P^i})$

$Prob_model(P^i, W^{P^i})$

3.1 Results based on the simulated data

DACL and DACR are the most similar pair among the six models included in our experiment, making them the hardest to differentiate from each other. In Thorne and Stump[17], networks generated by DACR were used as the target to see if they could be detected correctly or would be mistaken as DACL. Here we did the same test for comparison. In addition, we also did the test in the reversed direction, namely, using the networks generated by DACL as the target to see how well ABC-DEP can detect and differentiate them from DACR and other models.

3.1.1 Data simulated by DACR—The first test data is simulated by DACR, with parameters $\delta = 0.4$, $\alpha = 0.25$, $p = 0.7$ and $m = 3$, grown to 5000 nodes with 25009 edges. The posterior model probabilities illustrated in Fig.4 show that the DACR has the highest average probability, which means ABC-DEP can accurately predict the model that is used to generate the test data. From the box-plot in Fig.5(a)(1), we can see that DACR's posterior probabilities converge rapidly and smoothly towards their expected values respectively. Compared to the traditional method ABC-SMC[17] which mistakenly predicted DACL as the underlying model for networks generated by DACR, our method is obviously more accurate. Moreover, for the parameter estimation of DACR shown by Fig.5(b), where the blue line means the average value of our estimation and the red line means the gold standard value used to generate the test network, the parameter distribution, although not very smooth, is almost centered around the correct values. The standard deviations for the estimation of δ, α, p and m are 0.0196, 0.0300, 0.1114 and 0.0700 respectively. And the range of parameters' distribution of our method is narrower than that of ABC-SMC[17]. We believe increasing the number of particles may make the distribution smoother and the mean possibly closer to the actual value, though it will cause additional computational cost. We therefore strike a balance between the accuracy and computational complexity. Besides, we have reproduced the traditional method ABC-SMC[17] to make further comparison. Note that the posterior probability from ABC-SMC incorrectly converges towards DACL for test networks generated by DACR as illustrated in Fig.5(a)(2). In additional to posterior probability, we also analyzed the minimum distance, as measured by Eq.(5), between the simulation network and the observed test network, as shown in Fig.5(c). Our results show a better convergence; and also the minimum distance obtained by our method can reach a smaller value.

3.1.2 Data simulated by DACL—To further demonstrate the robustness and accuracy of our method, we tested our method by an additional simulated test data. This second test data is simulated by DACL, also with parameters $\delta = 0.4$, $\alpha = 0.25$, $p = 0.7$ and $m = 3$, and grown to 5000 nodes with 25009 edges. Fig.7 illustrates the posterior probabilities of six models, and Fig.6(a)(1) and (2) illustrate the converging processes of posterior probabilities and minimum distances respectively. And the histogram of parameter estimation of DACL shown by Fig.6(b) is quite smoothly distributed and centered around the correct values, with standard deviations 0.0270, 0.0364, 0.1866 and 0.0861 for the estimation of δ, α, p and m respectively.

3.2 Results based on protein interaction data

We applied ABC-DEP to high-confidence human PPI network and high-confidence yeast PPI network that were downloaded from PrePPI database[19], [20], where high-confidence means the interactions in the datasets are at least supported by two publications. As a quick summary, the high-confidence human PPI network has 4003 nodes and 6781 edges, and the high-confidence yeast PPI network has 3236 nodes and 11381 edges. The results illustrated in Fig.9 shows the DA is the predominant evolutionary mechanism for the high-confidence human PPI network, while SF is the predominant evolutionary mechanism for the high-confidence yeast PPI network. For these two models, the parameter estimations are shown by Fig.8(a) and Fig.8(b) respectively. The parameter estimation of *w* shown by Fig.8(b) is presumed > 0, that is why the distribution looks a little different. Generally, the center-around parameter distribution can help researchers to choose most suitable parameters for the evolutionary models. While scale-free is an important topology property of PPI network[9], [27], not all scale-free PPI networks are born equal[28]. Considering that our graph spectra based method can capture many aspects of network structure, so the simulation networks generated by different evolutionary models may be scale-free, yet different in other topological properties such as betweenness, modularity, clustering coefficient and so on, a phenomenon reported in recent literature[11]. In this regard, our finding of DA as the predominant mechanism for high-confidence human PPI network is consistent with other recent studies[29]. There is no consensus currently with respect to topological characteristics of yeast PPI networks; while some reported scale-free[30], others did not[31]. The importance of our method is to help researchers analyze PPI network comprehensively from the evolutionary perspective, and provide a reference for them to find a possible evolutionary model given existing PPI networks. With PPI networks is becoming more and more complete, we believe our method will become increasingly important.

4 Conclusions

We have developed a novel model selection and parameter estimation method, ABC-DEP, based on Approximate Bayesian Computation and modified Differential Evolution. The results based on simulated data illustrate the efficacy of ABC-DEP. Detailed comparisons between our method and T. Thorne and M. P. H. Stumpf[17] have been made, which shows ABC-DEP has competitive advantages in accuracy and efficiency. Furthermore, we applied our method to real PPI networks data from human and yeast. The results show that DA model is the predominant evolutionary mechanism for the high-confidence human PPI

network, and SF model as the predominant evolutionary mechanism for the high-confidence yeast PPI network. Given the strong performance on the simulated data, we believe that our method provides a very useful tool for researchers to select and develop PPI evolutionary models and may also help resolve controversy regarding topological characteristics of PPI networks from the evolutionary perspective. Our method is highly parallelizable and it is our plan to pursue faster parallel implementation to meet the computational demands as PPI networks scale up.

Acknowledgments

Funding: Delaware INBRE program, with grants from the National Center for Research ResourcesNCRR (5P20RR016472-12) and the National Institute of General Medical SciencesNIGMS (8 P20 GM103446-12) from the National Institutes of Health.

References

1. Ispolatov I, Krapivsky P, Mazo I, Yuryev A. Cliques and duplication-divergence network growth. New J Phys. 2005; 7:145. [Online]. Available: http://www.ncbi.nlm.nih.gov/pubmed/18239727.
2. Ispolatov I, Krapivsky PL, Yuryev A. Duplication-divergence model of protein interaction network. Phys Rev E Stat Nonlin Soft Matter Phys. 2005; 71(6 Pt 1):061911. [Online]. Available: http://www.ncbi.nlm.nih.gov/pubmed/16089769. [PubMed: 16089769]
3. Peterson GJ, Presse S, Peterson KS, Dill KA. Simulated evolution of protein-protein interaction networks with realistic topology. PLoS One. 2012; 7(6):e39052. [Online]. Available: http://www.ncbi.nlm.nih.gov/pubmed/22768057. [PubMed: 22768057]
4. A-L, AR Barabási. Emergence of scaling in random networks. Science. 1999; 286(5439):509–512. [Online]. Available: http://www.sciencemag.org/content/286/5439/509.abstract. [PubMed: 10521342]
5. Brown KS, Hill CC, Calero GA, Myers CR, Lee KH, Sethna JP, Cerione RA. The statistical mechanics of complex signaling networks: nerve growth factor signaling. Phys Biol. 2004; 1(3–4): 184–195. [Online]. Available: http://www.ncbi.nlm.nih.gov/pubmed/16204838. [PubMed: 16204838]
6. Han JD, Bertin N, Hao T, Goldberg DS, Berriz GF, Zhang LV, Dupuy D, Walhout AJ, Cusick ME, Roth FP, Vidal M. Evidence for dynamically organized modularity in the yeast protein-protein interaction network. Nature. 2004; 430(6995):88–93. [Online]. Available: http://www.ncbi.nlm.nih.gov/pubmed/15190252. [PubMed: 15190252]
7. Khanin R, Wit E. How scale-free are biological networks. Journal of Computational Biology. 2006; 13(3):810–818. [Online]. Available: ⟨GotoISI⟩://WOS:000237966000014. [PubMed: 16706727]
8. Tanaka R, Yi TM, Doyle J. Some protein interaction data do not exhibit power law statistics. Febs Letters. 2005; 579(23):5140–5144. [Online]. Available: ⟨GotoISI⟩://WOS:000232194300007. [PubMed: 16143331]
9. Jeong H, Mason SP, Barabasi AL, Oltvai ZN. Lethality and centrality in protein networks. Nature. 2001; 411(6833):41–42. [Online]. Available: http://www.ncbi.nlm.nih.gov/pubmed/11333967. [PubMed: 11333967]
10. Patil A, Kinoshita K, Nakamura H. Domain distribution and intrinsic disorder in hubs in the human proteinprotein interaction network. Protein Science. 2010; 19(8):1461–1468. [Online]. Available: http://dx.doi.org/10.1002/pro.425. [PubMed: 20509167]
11. Joy MP, Brock A, Ingber DE, Huang S. High-betweenness proteins in the yeast protein interaction network. J Biomed Biotechnol. 2005; 2005(2):96–103. [Online]. Available: http://www.ncbi.nlm.nih.gov/pubmed/16046814. [PubMed: 16046814]
12. S-H Y, ZN O, A-L B. Functional and topological characterization of protein interaction networks. PROTEOMICS. 2004; 4(4):928–942. [Online]. Available: http://dx.doi.org/10.1002/pmic.200300636. [PubMed: 15048975]

13. Travers J, Milgram S. An experimental study of the small world problem. Sociometry. 1969; 32(4): 425–443. [Online]. Available: http://links.jstor.org/sici?sici=0038-0431%28196912%2932%3A4%3C425%3AAESOTS%3E2.0.CO%3B2-W.
14. Ravasz E, Somera AL, Mongru DA, Oltvai ZN, Barabasi AL. Hierarchical organization of modularity in metabolic networks. Science. 2002; 297(5586):1551–1555. [Online]. Available: http://www.ncbi.nlm.nih.gov/pubmed/12202830. [PubMed: 12202830]
15. Maslov S, Sneppen K. Specificity and stability in topology of protein networks. Science. 2002; 296(5569):910–913. [Online]. Available: http://www.ncbi.nlm.nih.gov/pubmed/11988575. [PubMed: 11988575]
16. Wan X, Cai S, Zhou J, Liu Z. modularity and disassortativity in protein-protein interaction networks. Chaos. 2010; 20(4):045113. [Online]. Available: http://www.ncbi.nlm.nih.gov/pubmed/21198125. [PubMed: 21198125]
17. Thorne T, Stumpf MP. Graph spectral analysis of protein interaction network evolution. J R Soc Interface. 2012; 9(75):2653–2666. [Online]. Available: http://www.ncbi.nlm.nih.gov/pubmed/22552917. [PubMed: 22552917]
18. Rainer S, Kenneth P. Differential evolution – a simple and efficient heuristic for global optimization over continuous spaces. J of Global Optimization. 1997; 11(4):341–359.
19. Zhang QC, Petrey D, Deng L, Qiang L, Shi Y, Thu CA, Bisikirska B, Lefebvre C, Accili D, Hunter T, Maniatis T, Califano A, Honig B. Structure-based prediction of protein-protein interactions on a genome-wide scale. Nature. 2012; 490(7421):556–560. [PubMed: 23023127]
20. Zhang QC, Petrey D, Garzon JI, Deng L, Honig B. Preppi: a structure-informed database of protein-protein interactions. Nucleic Acids Res. 2013; 41(Database issue):D828–D833. [Online]. Available: http://www.ncbi.nlm.nih.gov/pubmed/23193263. [PubMed: 23193263]
21. Csillery K, Blum MG, Gaggiotti OE, Francois O. Approximate bayesian computation (abc) in practice. Trends Ecol Evol. 2010; 25(7):410–418. [Online]. Available: http://www.ncbi.nlm.nih.gov/pubmed/20488578. [PubMed: 20488578]
22. Umeyama S. An eigendecomposition approach to weighted graph matching problems. Pattern Analysis and Machine Intelligence, IEEE Transactions on. 1988; 10(5):695–703.
23. Wilson RC, Zhu P. A study of graph spectra for comparing graphs and trees. Pattern Recognition. 2008; 41(9):2833–2841. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0031320308000927.
24. Differential evolution for continuous function optimization (an algorithm by kenneth price and rainer storn). [Online]. Available: http://www1.icsi.berkeley.edu/n_storn/code.html.
25. Toni T, Stumpf MP. Simulation-based model selection for dynamical systems in systems and population biology. Bioinformatics. 2010; 26(1):104–110. [Online]. Available: http://www.ncbi.nlm.nih.gov/pubmed/19880371. [PubMed: 19880371]
26. Dorogovtsev S, Mendes J. Minimal models of weighted scale-free networks. 2004 [Online]. Available: http://arxiv.org/abs/cond-mat/0408343.
27. Barabási AL, Oltvai ZN. Network biology: understanding the cell's functional organization. Nat Rev Genet. 2004; 5(2):101–113. [Online]. Available: http://www.ncbi.nlm.nih.gov/pubmed/14735121. [PubMed: 14735121]
28. Hormozdiari F, Berenbrink P, Przulj N, Sahinalp SC. Not all scale-free networks are born equal: the role of the seed graph in ppi network evolution. PLoS Comput Biol. 2007; 3(7):e118. [Online]. Available: http://www.ncbi.nlm.nih.gov/pubmed/17616981. [PubMed: 17616981]
29. Jin Y, Turaev D, Weinmaier T, Rattei T, Makse HA. The evolutionary dynamics of protein-protein interaction networks inferred from the reconstruction of ancient networks. PLoS One. 2013; 8(3):e58134. [Online]. Available: http://www.ncbi.nlm.nih.gov/pubmed/23526967. [PubMed: 23526967]
30. Albert R. Scale-free networks in cell biology. J Cell Sci. 2005; 118(Pt 21):4947–4957. [Online]. Available: http://www.ncbi.nlm.nih.gov/pubmed/16254242. [PubMed: 16254242]
31. Hase T, Tanaka H, Suzuki Y, Nakagawa S, Kitano H. Structure of protein interaction networks and their implications on drug design. Plos Computational Biology. 2009; 5(10) [Online]. Available: ⟨GotoISI⟩://WOS:000272033100030.

Biographies

Lei Huang is a PhD student in Department of Computer and Information Sciences, University of Delaware. His research interests focus on bioinformatics and maching learning, specifically on evolutionary analysis and interaction prediction for protein-protein interaction network.

Li Liao received the PhD degree in theoretical physics from Peking University in 1992. Currently, he is working as an associate professor of computer and information sciences at the University of Delaware. He has worked in the field of bioinformatics for about 15 years, with broad experience in developing computational methods such as hidden Markov models and support vector machines to solve a wide variety of biological problems. He is active in research and serving the bioinformatics community and also served as program committee member and/or organizer for more than 20 conferences and workshops in bioinformatics for the past 5 years, and is currently on the editorial board of two journals. He is a senior member of the ACM and an author or coauthor of 50 peer-reviewed publications.

Cathy H. Wu is the Edward G. Jefferson chair and director of the Center for Bioinformatics and Computational Biology, professor of the Departments of Computer and Information Sciences and of Biological Sciences, and director of the Bioinformatics Masters Program at the University of Delaware. She is also an adjunct professor of biochemistry and molecular biology at Georgetown University. With educational background in both biology and computer science, she has conducted bioinformatics research for 20 years. She has directed the Protein Information Resource (PIR) since 2001 as a major bioinformatics resource to support genomics, proteomics, and systems biology research. The PIR websites are accessible by researchers worldwide with more than 30 million hits per month from over 100,000 unique sites. Her research encompasses protein structure-function evolution, biological text mining and ontology, proteomic informatics, computational systems biology, and translational bioinformatics. She is the PI/Co-PI on several large consortium projects, including the UniProt, Protein Ontology, and BioCreative. She serves on several advisory boards, including the NIGMS Protein Structure Initiative, the Association for Computing

Machinery (ACM) SIGBioinformatics, and US Human Proteome Organization (HUPO), and has served on more than 40 conference organizing or scientific committees. She has published six books and conference proceedings and more than 160 peer-reviewed papers, and given more than 120 invited talks.

Fig. 1.
Flowchart of DE algorithm[24].

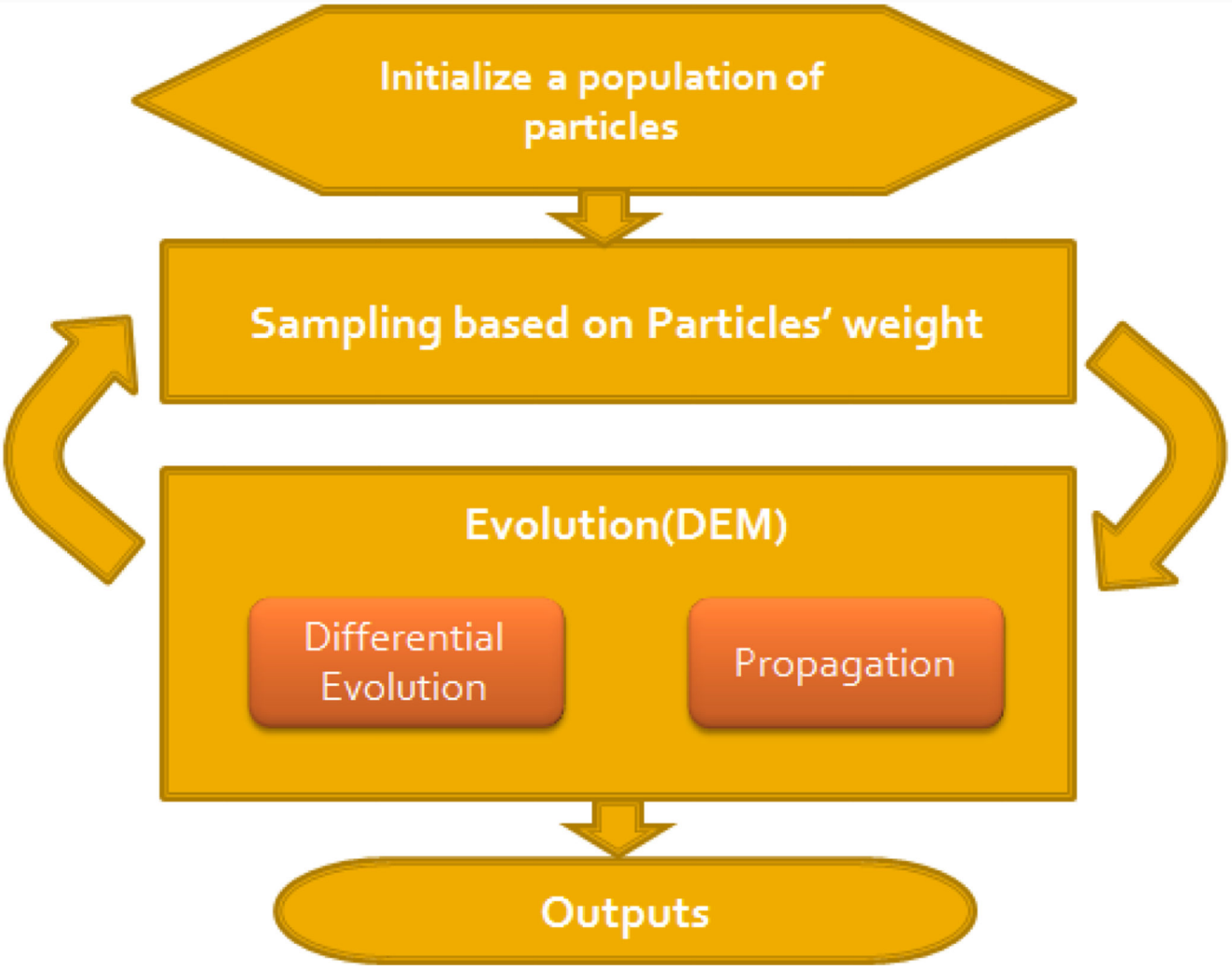


Fig. 2.
ABC-DEP process.

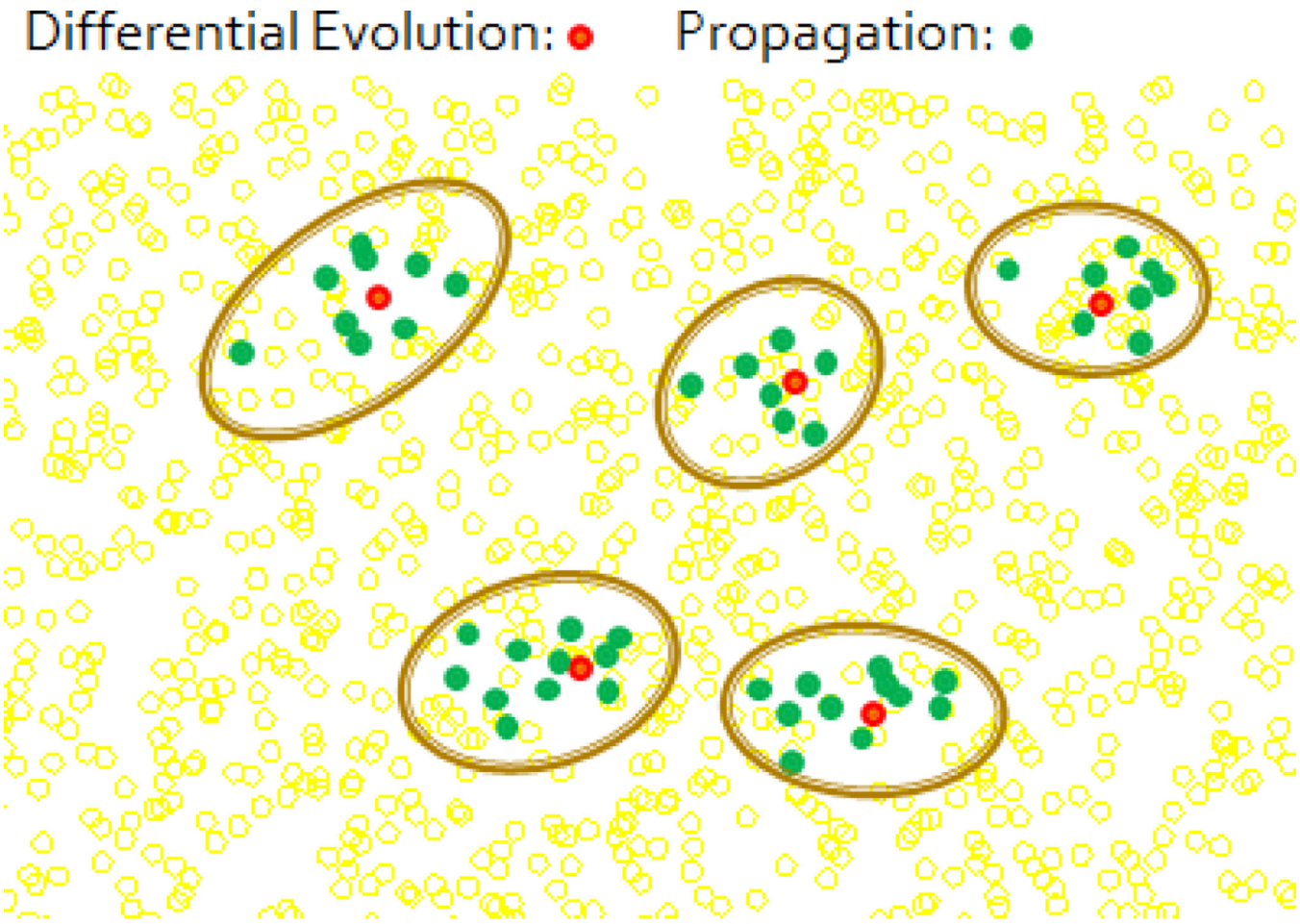


Fig. 3.
Functions of DE and Propagation.

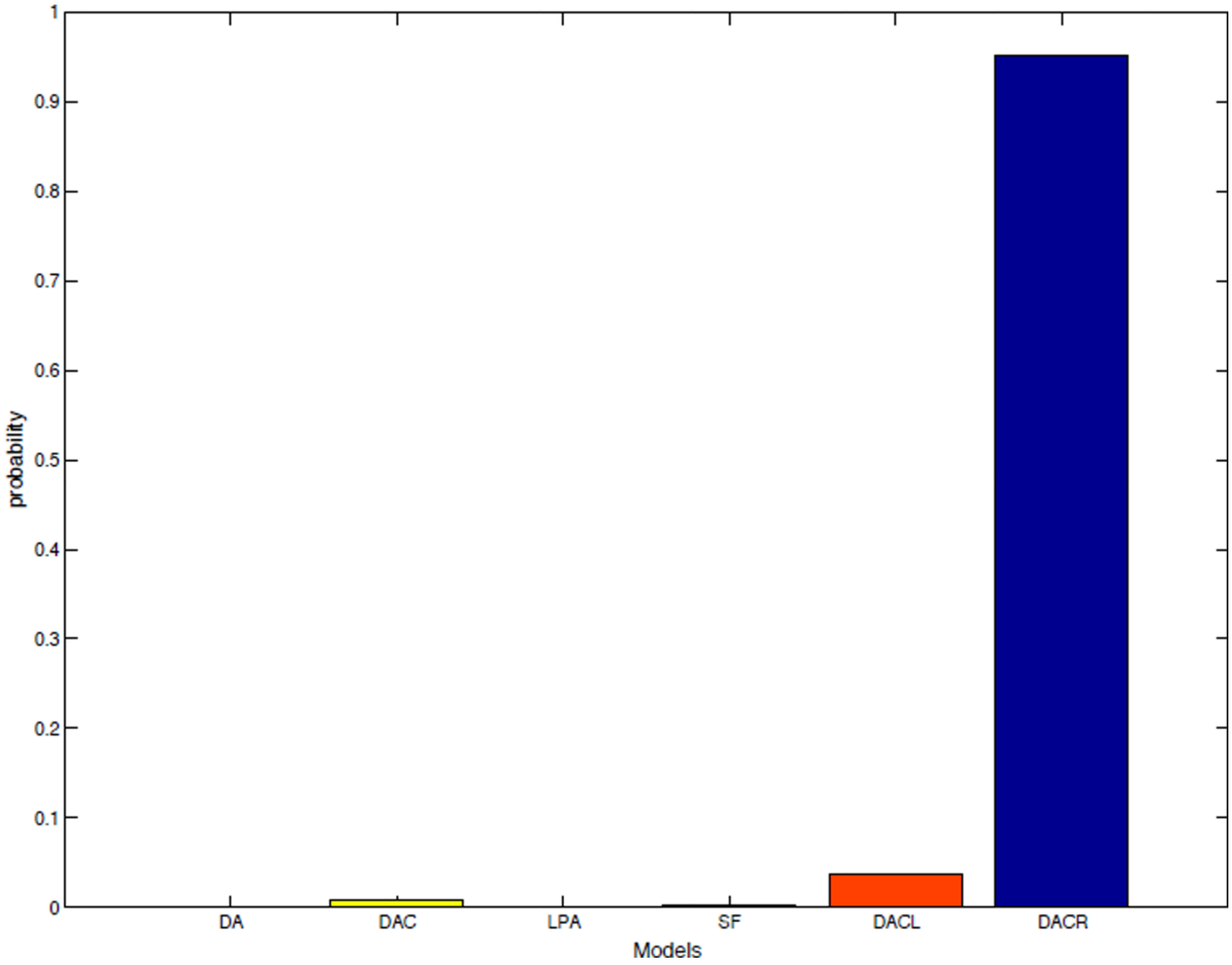
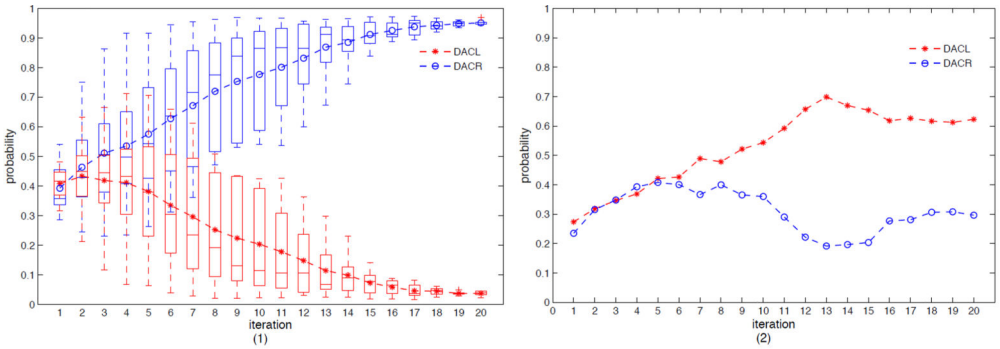
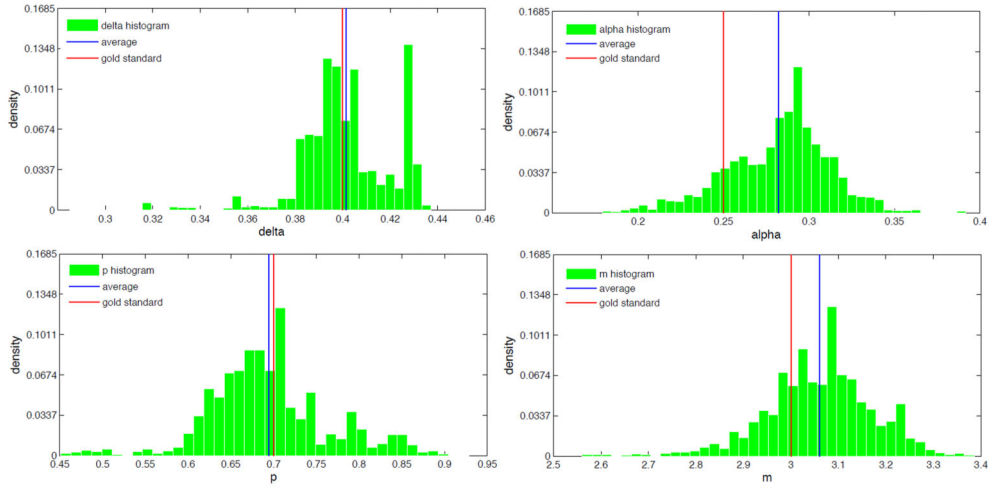


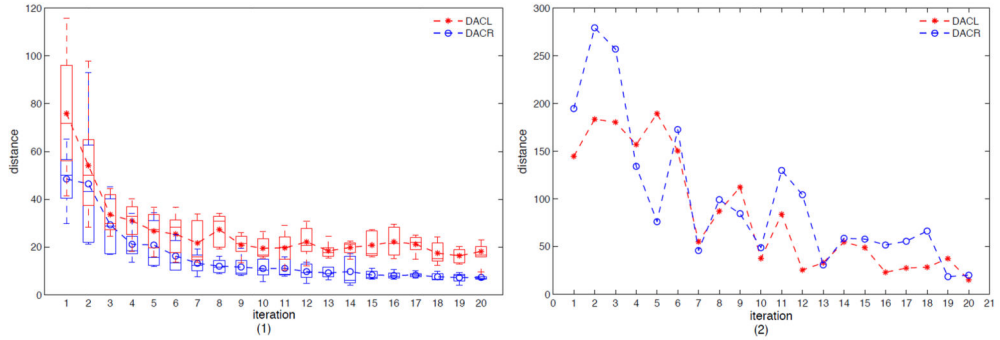
Fig. 4.
Posterior probabilities based on test data simulated by DACR.



(a) Comparison of posterior probability converging process between our method(1) and ABC-SMC(2)[17]

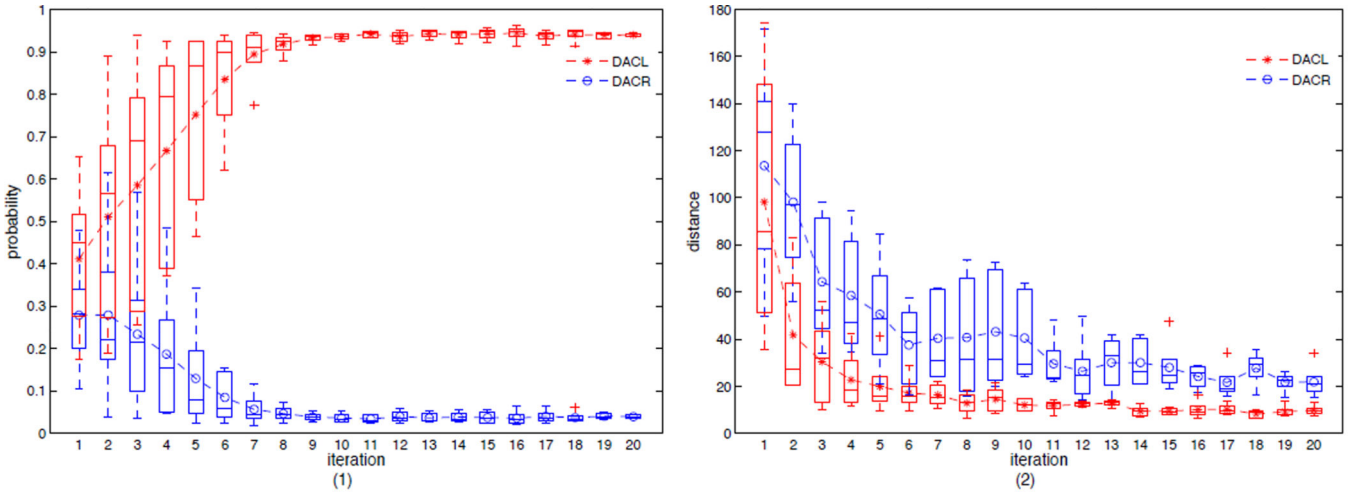


(b) Parameter estimation based on our method

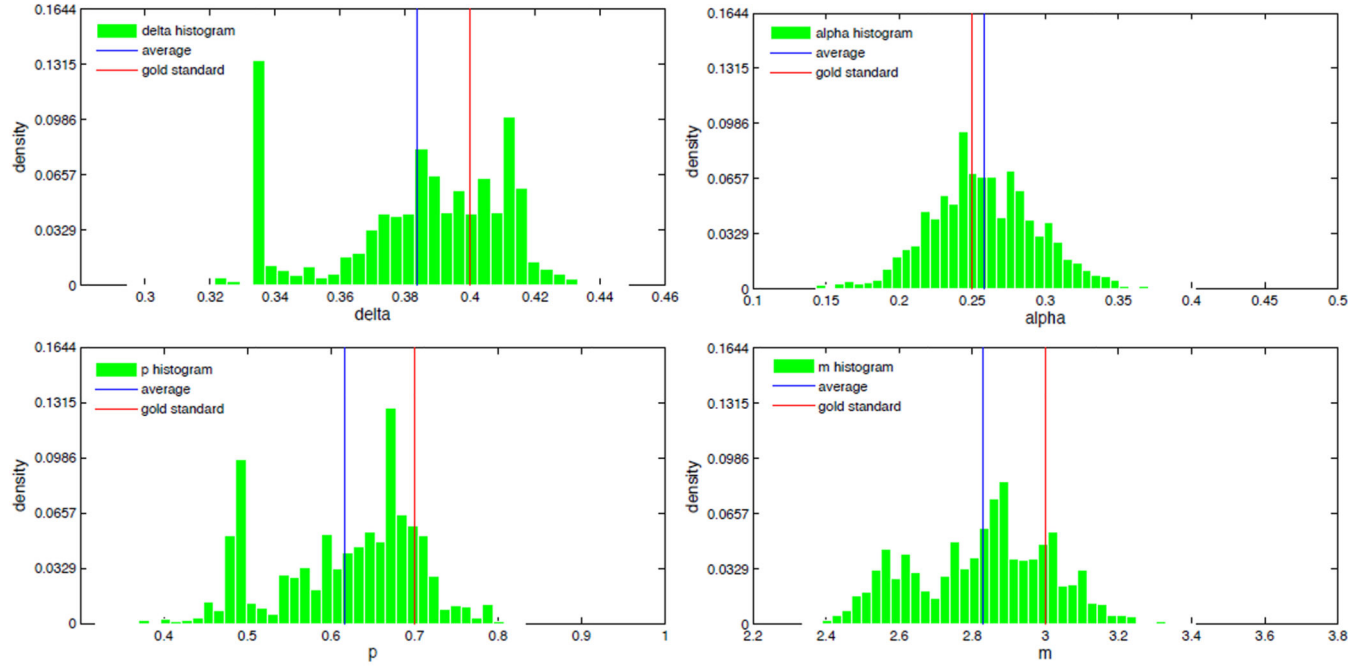


(c) Comparison of minimum distance converging process between our method(1) and ABC-SMC(2)[17]

Fig. 5.
Results based on testing data simulated by DACR



(a) Coverging process of porsterior probability(1) and minimum distance(2)



(b) Parameter estimation based on our method

Fig. 6.
Results based on testing data simulated by DACL

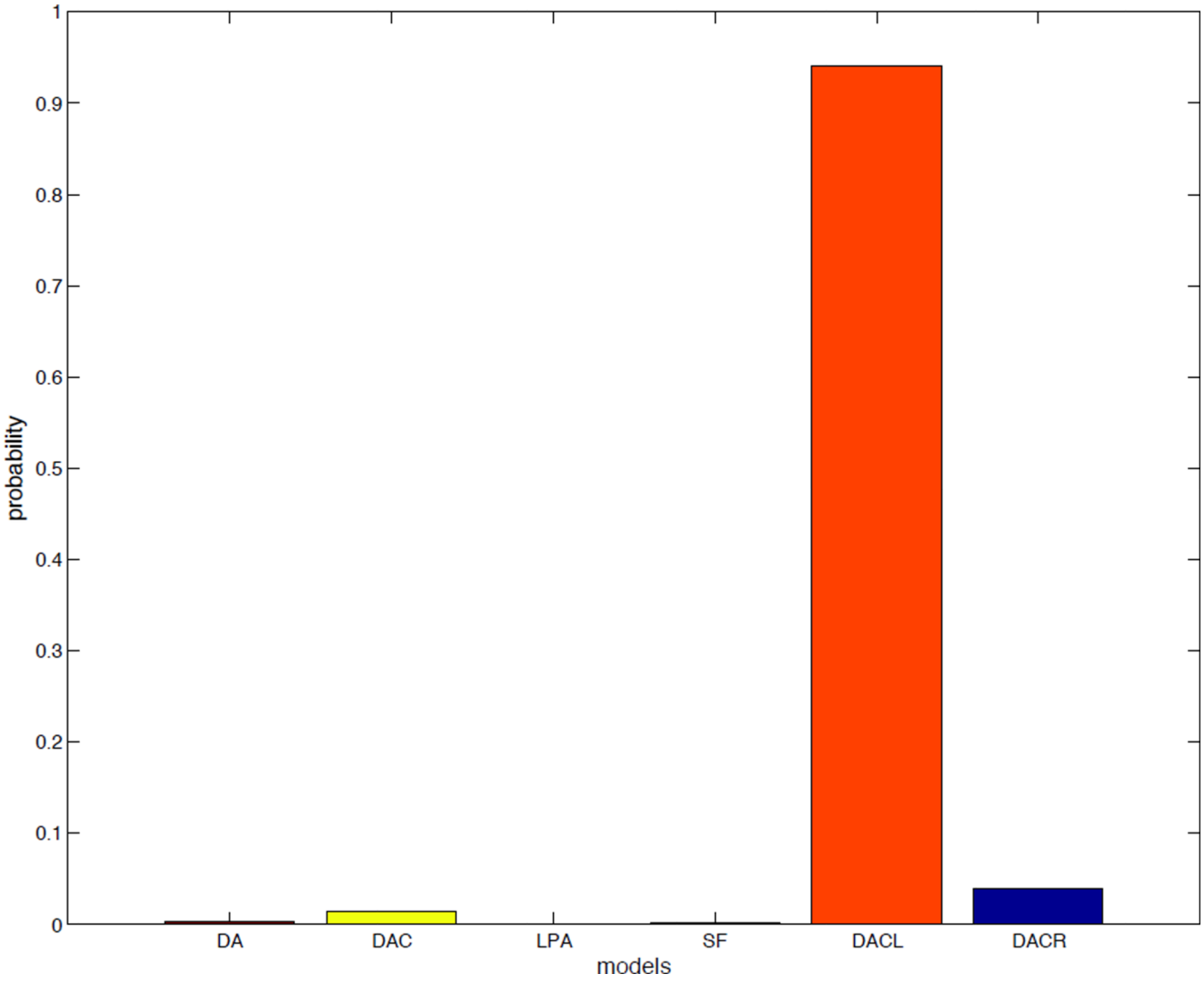


Fig. 7.
Posterior probabilities based on test data simulated by DACL.

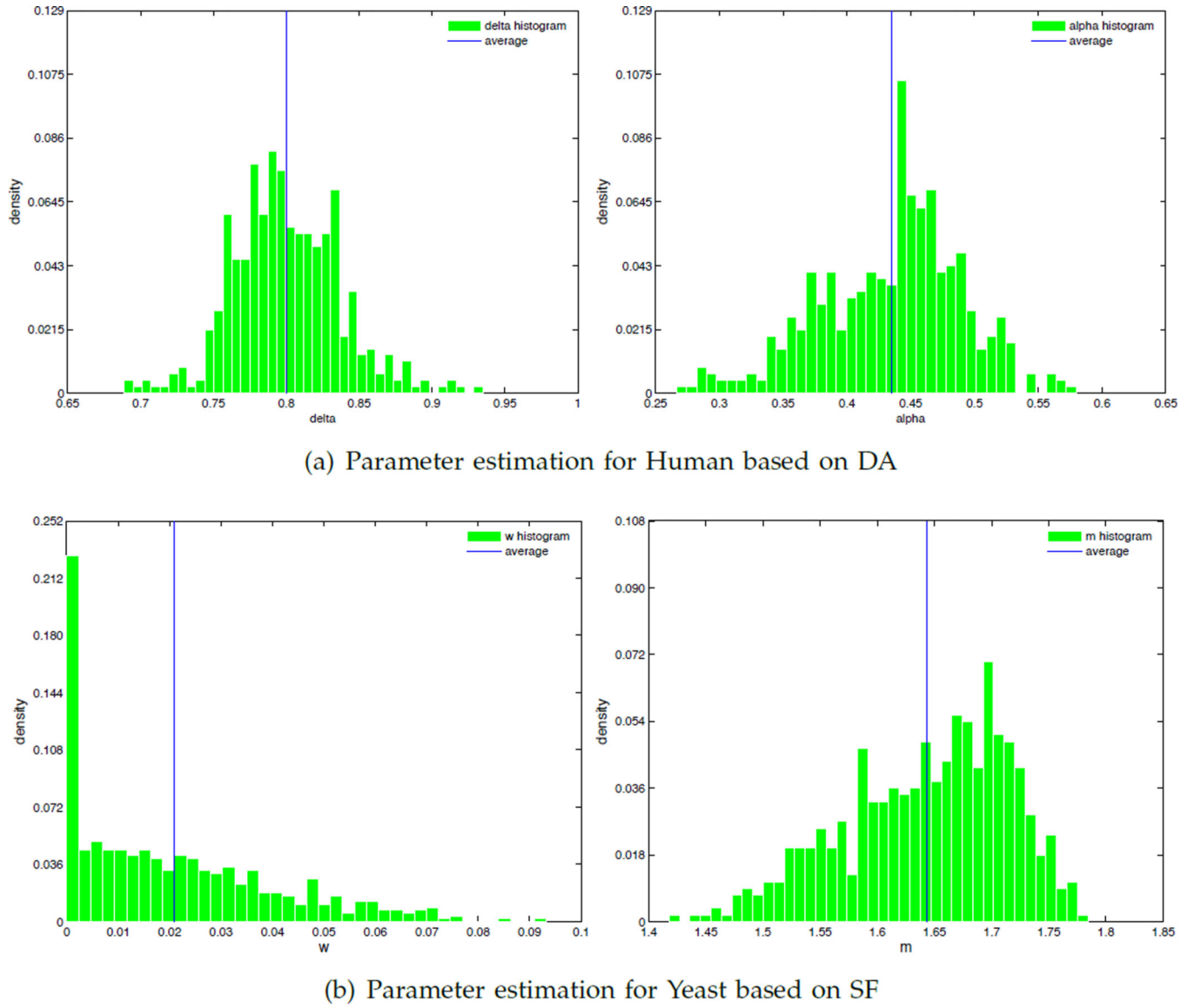


Fig. 8.
Parameters estimation for Human and Yeast based on their preferred models

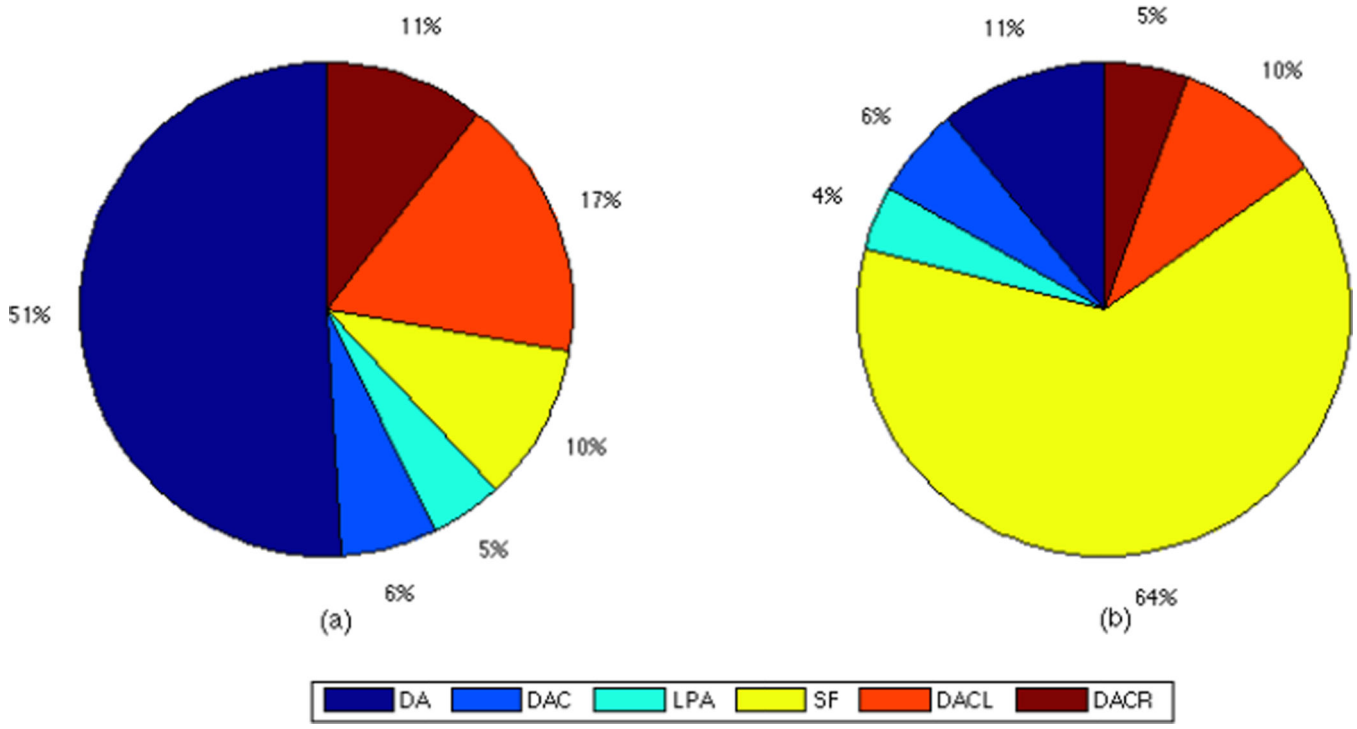


Fig. 9.
Evolutionary model prediction for real PPI networks.(a) is for Human, (b) is for Yeast