Examination of Higher Order Epistasis and Fitness Peaks through Analysis of 4-Locus Fitness Graphs

Mengming Luo

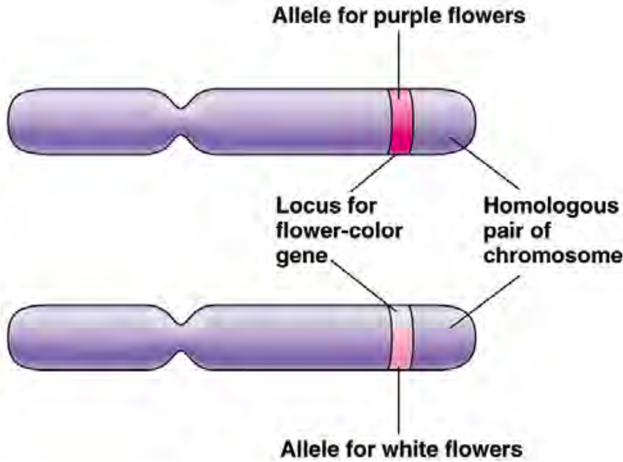
February 1, 2019

Outline

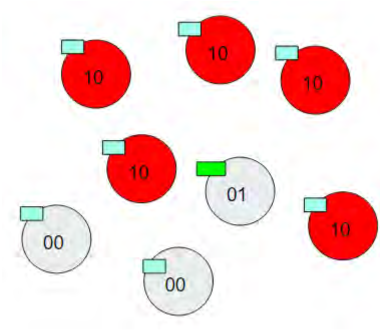
- Background
- Goal
- Applications/Importance
- Theorems
- Results and Conclusions

Basic Terminology

- Loci
- Alleles
- Genotype
- Phenotype

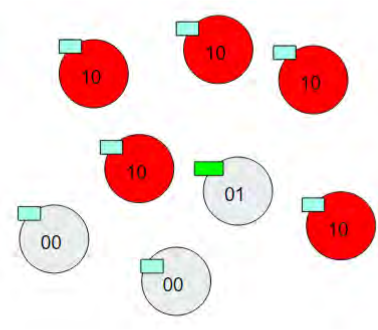


Terminology



- Population
- Biallelic
  1. Wild type allele $\rightarrow 0$
  2. Mutational allele $\rightarrow 1$
- N-loci system
  1. 2-loci: Wild Type (00), Single Mutant (01, 10), Double Mutant (11)
- Mutational neighbors
- Example 2-locus system
  1. Locus A – circle color
  2. Locus B – rectangle color

Terminology

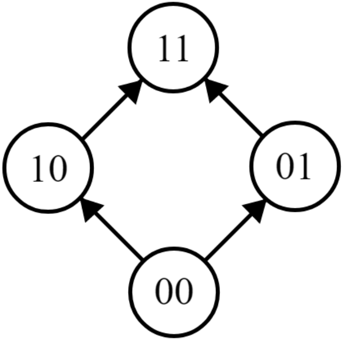


- Example 2-locus system
  1. Locus A – circle color
  2. Locus B – rectangle color
- Set of alleles
  1. $\Sigma = \{0, 1\}$
- Genotype space
  1. $\Sigma^2 = \{00, 01, 10, 11\}$.
- Fitness
  1. Expected reproductive success of genotype g

Fitness Landscapes and Graphs

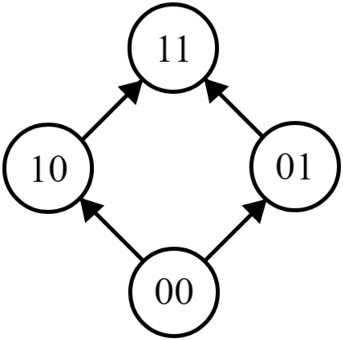
- Quantitative tools
- Fitness landscapes
  1. Function $w : \Sigma^n \to \mathbb{R}$
  2. Assigns a fitness value w_g to each genotype g in Σ^n
- Fitness Graphs
  1. Visual representations of fitness landscapes

Example 2-locus graph



- Mutational neighbors have arrows between them, representing evolutionary pathways
- The arrowhead points to the genotype with greater fitness.
- Peaks
- This example is ‘additive’

Partial Orders



- An order that does not include relations between every genotype
- Example:
  1. $w_{11} > w_{10} > w_{00}$
  2. $w_{11} > w_{01} > w_{00}$
- Offers no information on w_{10}'s relation to w_{01}.
- Rank orders offer relations between every genotype

Epistasis

- Occurs when one gene's expression is affected by interaction with a gene at another locus
- The fitness of a multiple mutant is unexpected given the fitness of the wild type and all single and double mutants.
- example: good + good = bad
- Implied epistasis
- Higher order epistasis
  1. Pairwise epistasis
  2. n-way epistasis

Goal

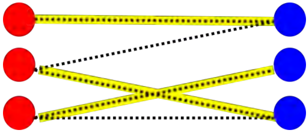
- Develop a characterization of n-locus graphs implying higher-order epistasis
- Apply the characterization to various 4-locus graphs and examine the relationship between number of peaks and higher-order epistasis

Applications/Importance

- Antibiotic Resistance
- Occurs when population of microbes evolves to resist an antibiotic that previously worked
- Demand for new antibiotics, higher medical costs, and higher rates of mortality
- Recent treatment plans use quantitative tools such as fitness graphs to approach the design of drug-cycling trials

Bipartite Graphs

- $G = (V, E)$ is bipartite if $E \neq \emptyset$ and if V can be partitioned into two disjoint subsets A and B, i.e., $V = A \cup B$ and $A \cap B = \emptyset$, such that for every edge $e \in E$, $e = (v_1, v_2)$, where $v_1 \in A$ and $v_2 \in B$
- Perfect matching – A subset of the edges that pairs every node in A with a different node in B.



Hall's Marriage Theorem

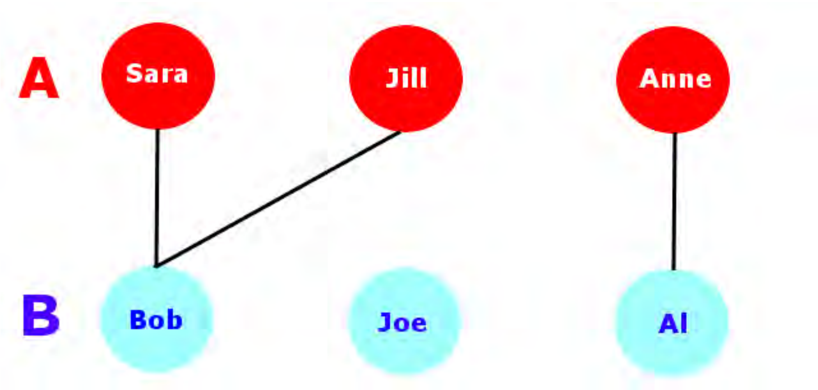
Given a finite bipartite graph $G = (V, E)$ with the bipartition $V = A \cup B, A \cap B = \emptyset$, $|A| = |B|$,

the graph G has a perfect matching if and only if

for every subset $S \subseteq A$, $|N(S)| \geq |S|$, where

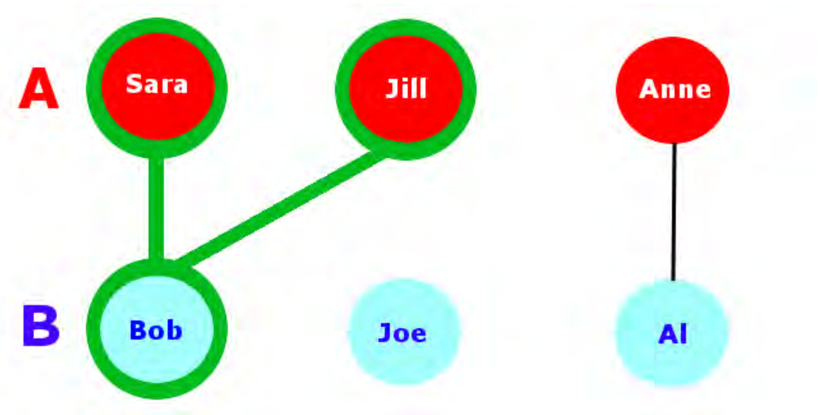
$$N(S) = \{b \in B, \text{ there exists } a \in S \text{ with } (a, b) \in E\}.$$

Illustration: Hall's Marriage Theorem



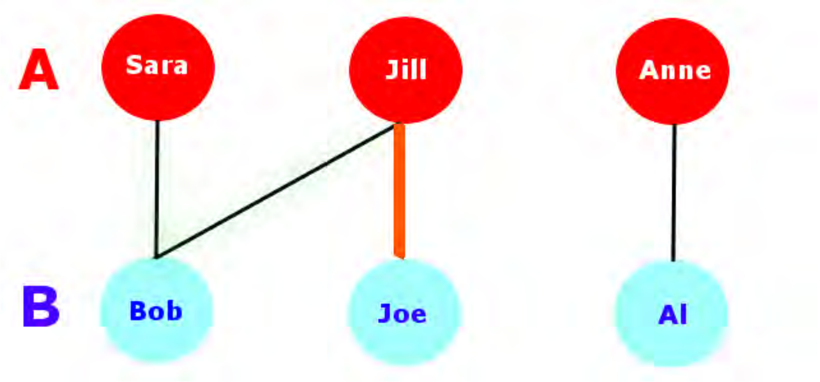
Bipartition of women and men. Edges connect women to men they want to marry.

Illustration: Hall's Marriage Theorem



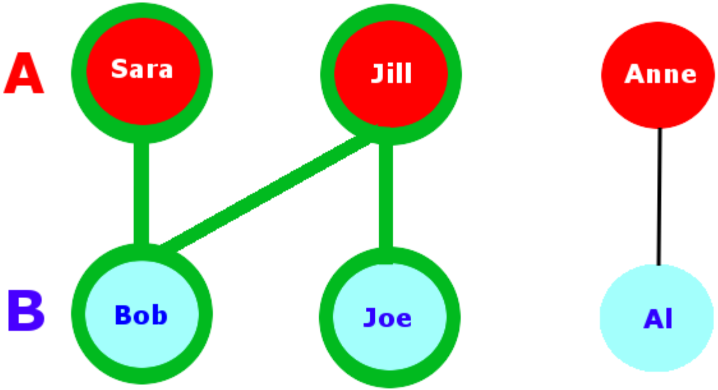
S = {Sara, Jill}, N(s) = {Bob}. $|N(S)| < |S|$

Illustration: Hall's Marriage Theorem



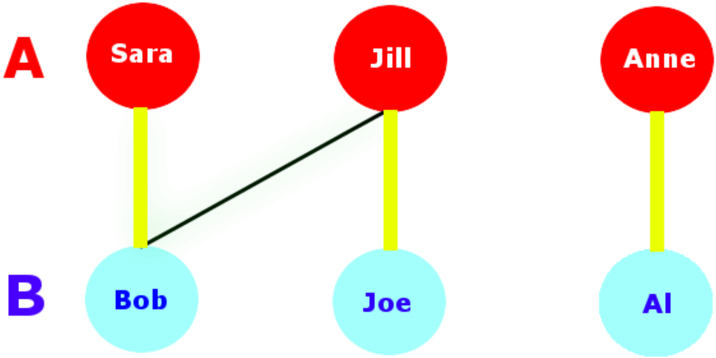
Add new edge to increase $|N(S)|$

Illustration: Hall's Marriage Theorem



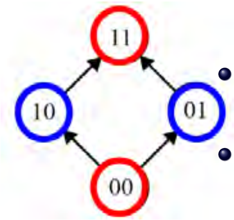
S = {Sara, Jill}, N(S) = {Bob, Joe}. $|N(S)| \geq |S|$

Illustration: Hall's Marriage Theorem



Perfect Matching = Happy Marriage!

Even/Odd Genotypes



- Even genotypes have an even number of '1's
  1. Ex: 11
- Odd genotypes have an odd number of '1's
  1. Ex: 10
- All nodes in a fitness graph are adjacent to only mutational neighbors
- Mutational neighbors differ by only one bit
- All even genotypes are mutations neighbors with only odd genotypes, and vice versa
- Thus all fitness graphs are bipartite – even/odd bipartition

Epistasis (ϵ)

- Epistasis is measured by deviance from additive fitness

$$\epsilon_n = \sum_{g \in \Sigma^n} (-1)^g w_g = \sum_{j_1=0}^{1} \sum_{j_2=0}^{1} \cdots \sum_{j_n=0}^{1} (-1)^{j_1+j_2+\cdots+j_n} w_{j_1 j_2 \ldots j_n}.$$

- For example, in a biallelic 2-locus system,

$$\epsilon_2 = w_{00} - w_{01} - w_{10} + w_{11}$$

.

- Note: In an additive 2-locus system,

$$(w_{11} - w_{00}) = (w_{01} - w_{00}) + (w_{10} - w_{00})$$

$$w_{11} - w_{00} = w_{01} + w_{10} - 2w_{00}$$

$$w_{11} - w_{00} - (w_{01} + w_{10} - 2w_{00}) = 0$$

$$w_{11} - w_{00} - w_{01} - w_{10} + 2w_{00} = 0$$

$$w_{11} + w_{00} - w_{01} - w_{10} = 0$$

Characterization

Partition Condition

A n-locus fitness graph G implies negative n-way epistasis

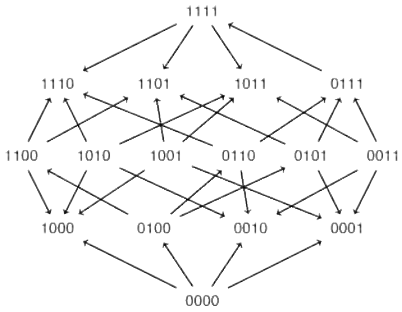
if and only if there exists a partition of 2^{n-1} disjoint pairs $\{o_j, e_j\}$ for all genotypes in Σ^n such that

$w_{o_j} > w_{e_j}$ for all $j = 1, 2, \cdots, 2^{n-1}$.

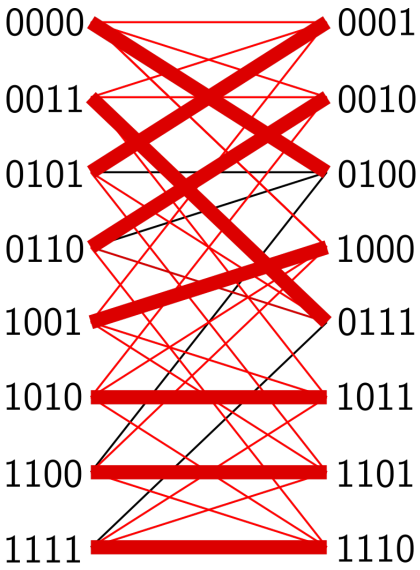
$$\epsilon_n = \sum_{g \in \Sigma^n} (-1)^g w_g = \sum_{j_1=0}^{1} \sum_{j_2=0}^{1} \cdots \sum_{j_n=0}^{1} (-1)^{j_1+j_2+\cdots+j_n} w_{j_1 j_2 \ldots j_n}.$$

Note all even genotypes have positive coefficients while all odd genotypes have negative coefficients.

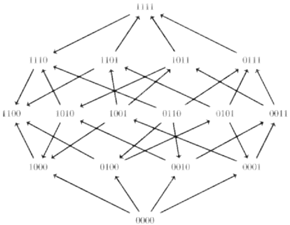
Characterization: Example - Implies Epistasis



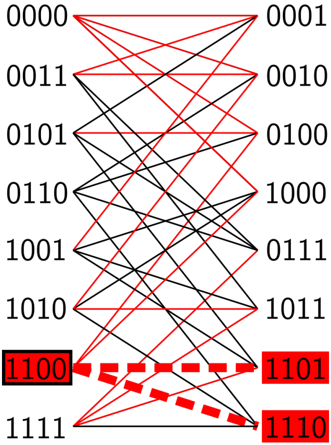
Edges where the odd genotype has greater fitness are in red. Bolded edges are part of the partition/perfect matching.



Characterization: Example - Does Not Imply Epistasis



Edges where the odd genotype has greater fitness are in red. Dashed lines: there is no partition/perfect matching where $w_{oj} > w_{ej}$ because there exists an $S_o = \{1101, 1110\} \subseteq V_o$ s.t. $|S_o| > |N(S_o)|$.



Results

# Peaks	# Graphs	2-way	3-way	4-way
8	1	0	0	1
7	4	0	1	3
6	59	1	10	48
6-8	64	1	11	52

# Peaks	4-way interactions
6 - 8	81%
5	50%
4	15%
3	10%
2	10%
1	5%
All graphs	9%

Conclusions

- A greater number of peaks is correlated with a higher chance of implying higher-order epistasis
- Evolving populations may end up “stuck” at a peak
  1. Multiple peaks, more likely to get stuck at a “suboptimal” peak, decreasing overall population fitness
- More peaks results in lower predictability

Further Study

- Application of characterization to other n-locus graphs
- Larger sample size for statistical analysis
- Practical analyses on populations for which there is little information on fitness rank orders
    1. Only information on partial orders
- Possible applications to data obtained from genome-wide association studies

Summary

- Developed a characterization of fitness graphs that imply higher-order epistasis based on Hall's Marriage Theorem
- Used that characterization to determine that higher order epistasis seems to correlate to the number of peaks in a graph

Acknowledgements

Thank you to...

Kristina Crona, mentor

Devin Greene, programmer

Ms. Bosse, SRP Teacher