

Swarm Intelligence Ant Colony Optimization

Based on slides by Thomas Bäck, which were based on:
Marco Dorigo and Thomas Stützle: Ant Colony Optimization. MIT Press,
Cambridge, MA, 2004.

Examples of Collective Intelligence in Nature



Termite hill



Nest of wasps



Flocking birds

Bee attack

Swarm Intelligence

- Originated from the study of colonies, or swarms of social organisms
- Collective intelligence arises from interactions among individuals having simple behavioral intelligence
- Each individual in a swarm behaves in a distributed way with a certain information exchange protocol

Communication

- **Point-to-point:** information between individuals or between an object and an individual is directly transferred
 - direct visual contact, antennation, trophallaxis (food or liquid exchange), chemical contact, ...
- **Broadcast-like:** the signal propagates to some limited extent throughout the environment and/or is made available for a rather short time
 - generic visual detection, use of lateral line in fishes to detect water waves, actual radio broadcast
- **Indirect (stigmergy):** two individuals interact indirectly when one of them modifies the environment and the other responds to the new environment at a later time
 - pheromone laying/following, post-it, web

Ant Colony Optimisation



© Bruce G. Marcot

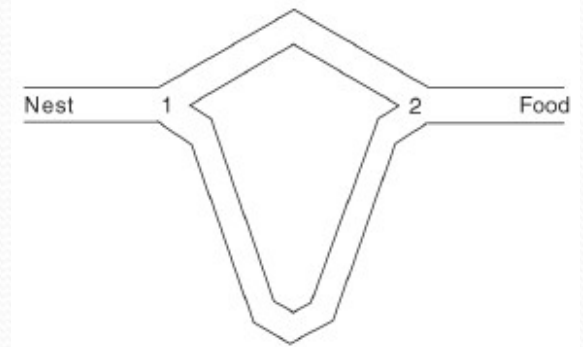
What is special about ants?

- Ants can perform complex tasks:
 - nest building, food storage
 - garbage collection, war
 - **foraging** (*to wander in search of food*)
- There is no management in an ant colony
 - collective intelligence
- They communicate using **pheromones** (*chemical substances*), sound, touch



Double Bridge Experiments

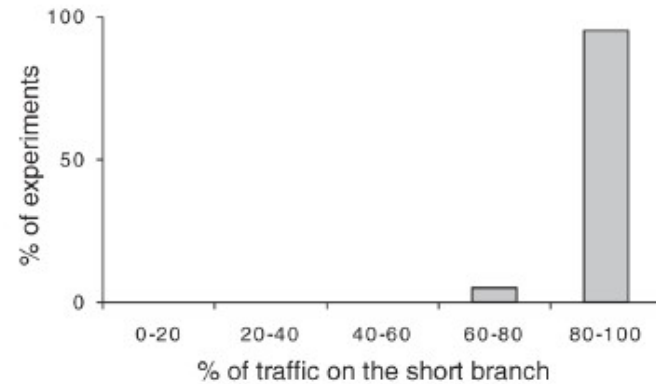
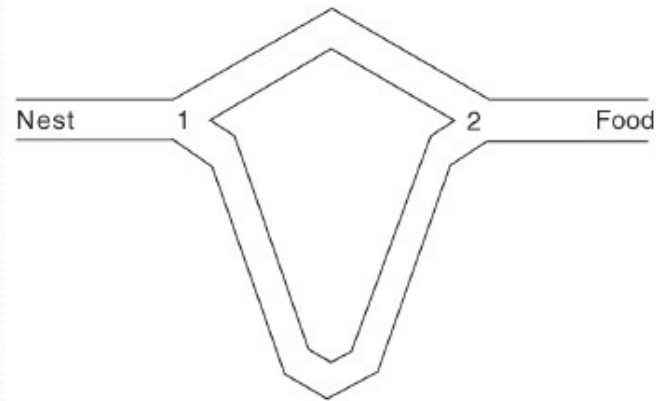
- A study on the pheromone trail-laying and -following behavior of Argentine ants
- A double bridge connects a nest of ants and a food source
- The ratio $r = L_{long} / L_{short}$ between the length of the two branches of the double bridge is varied
- Ants are free to move between the nest and the food



J.L. Deneubourg, S. Aron, S. Goss and J.M. Pasteels (1990). The self-organizing exploratory pattern of the Argentine ant. *Journal of Insect Behaviour*, 3, 159-168.

S. Goss, S. Aron, J.L. Deneubourg and J.M. Pasteels (1989). Self-organized shortcuts of the Argentine ant. *Naturwissenschaften*, 76, 579-581

Double Bridge Experiments



- In most of the trials, almost all the ants select the short branch (exploitation)
- Not all ants use the short branch, but a small percentage may take the longer one (exploration)

Foraging Behavior of Argentine Ants

- Ants initially explore the area surrounding their nest randomly
- Argentinian ants deposit pheromones everywhere they go
- When choosing their way, ants prefer to follow strong pheromone concentrations
- Pheromones defuse over time

Foraging Behavior of Argentine Ants

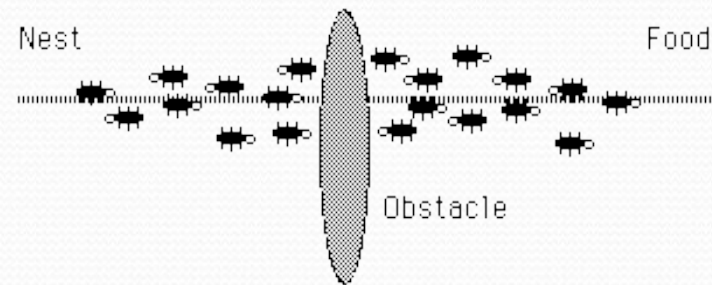
- How do Argentine ants find the shortest path?
 - The ants that take the shortest path arrive at the food source first
 - They return over the path that they took to get there, reinforcing the pheromones they deposited when going to the food source
 - Other ants notice the trail and follow it, reinforcing it further
- Hence, during the “start” of the experiment the advantage that ants on the shortest path had is reinforced

Alternative experiment

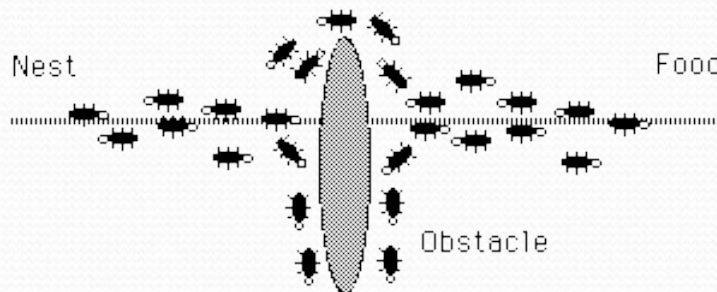
- An obstacle is put in the path of ants



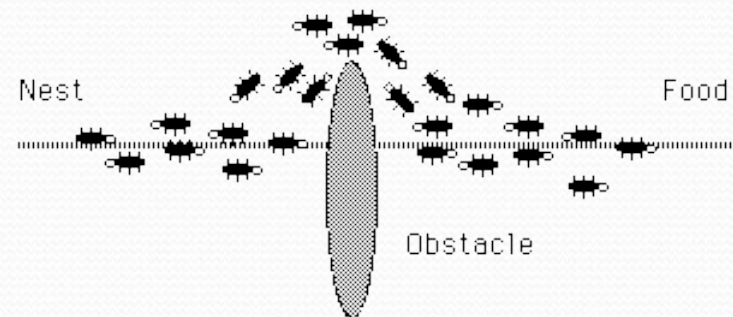
a) - Ants **follow path** between the Nest and the Food Source



b) - Ants go around the obstacle following one of two different paths with **equal probability**



c) Ants on the shortest path arrives at the food source first; on the way back they will follow the phero-mones on the shortest path again



d) – At the end, **all ants follow** the shortest path.

Simple Ant Colony Optimisation: Shortest Paths

- Artificial ants going “forward”
 - choose probabilistically the next node on their path, exploiting pheromones
 - do not drop pheromones
 - memorize the path they take
- Artificial ants going “backward”
 - deterministically follow the path they took earlier
 - drop pheromones proportionally to the quality of the path taken earlier

Simple ACO: Shortest Paths

```
initialize pheromones
for each iteration do
  for  $k = 1$  to number of ants do
    set out ant  $k$  at start node
    while ant  $k$  has not build a solution do
      choose the next node of the path
    end while
  end for
  update pheromones
end for
return best solution found
```

Simple ACO: Shortest Paths

- For an ant located at node v_i the probability p_{ij} of choosing v_j as the next node is:

$$p_{ij}^k = \begin{cases} \frac{(\tau_{ij})^\alpha}{\sum_{m \in N_i^k} (\tau_{im})^\alpha} & \text{if } j \in N_i^k \\ 0 & \text{if } j \notin N_i^k \end{cases}$$

where

- τ_{ij} is the amount of pheromones on edge $i \rightarrow j$
- N_i^k is the set of neighbors of node i not visited by ant k yet (tabu list)

Simple ACO: Shortest Paths

- Change in pheromone for an ant k on edge $i \rightarrow j$

$$\Delta\tau_{ij}^k = \begin{cases} Q/L_k & \text{if } (i, j) \in T_k \\ 0 & \text{otherwise} \end{cases}$$

where:

- Q : a heuristic parameter
- T_k : the path traversed by ant k
- L_k : the length of T_k calculated as the sum of all lengths of edges in T_k

Simple ACO: Shortest Paths

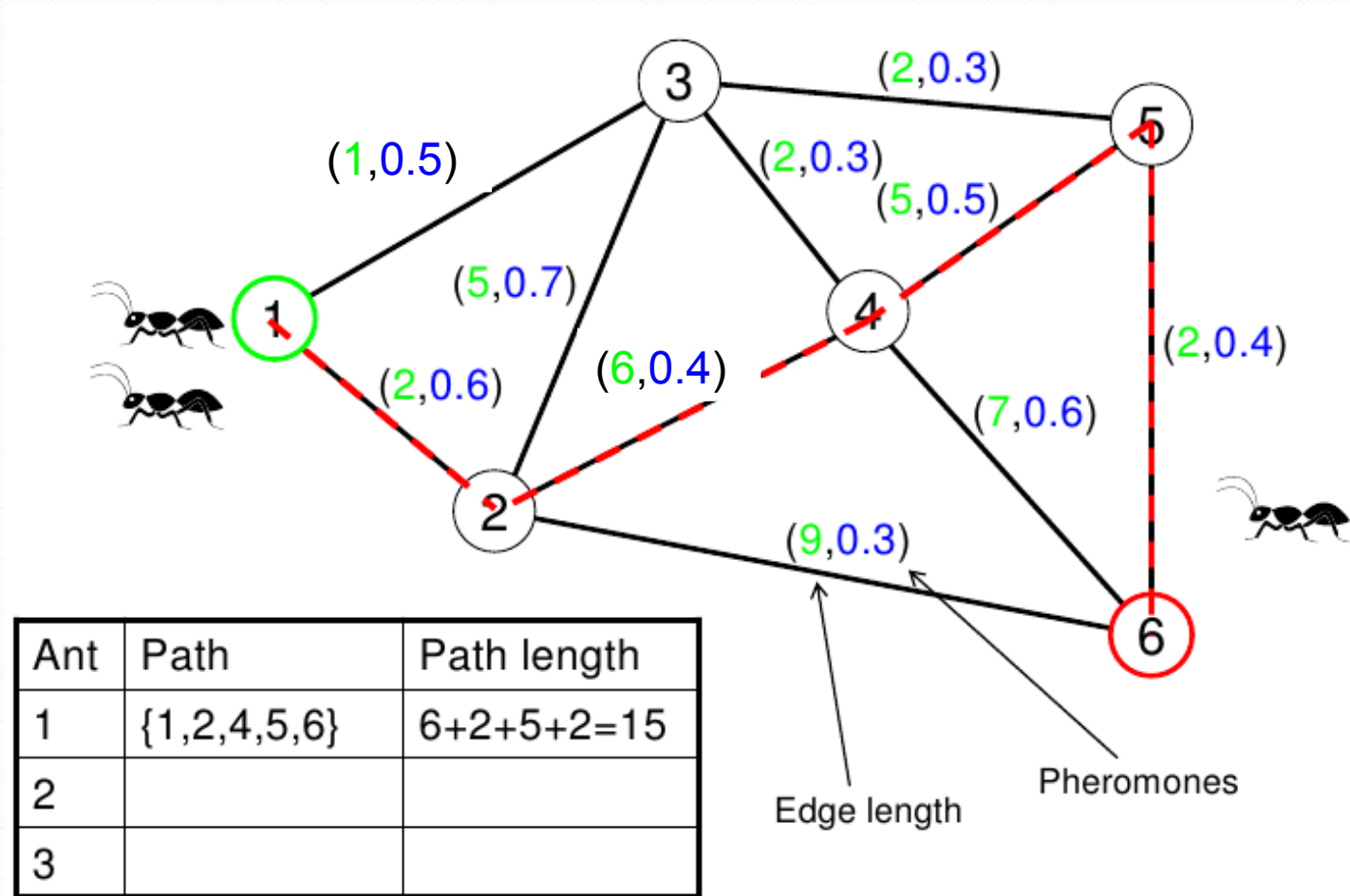
- Pheromone update on an edge $i \rightarrow j$

$$\tau_{ij} = (1 - \rho)\tau_{ij} + \sum_{k=0}^m \Delta\tau_{ij}^k$$

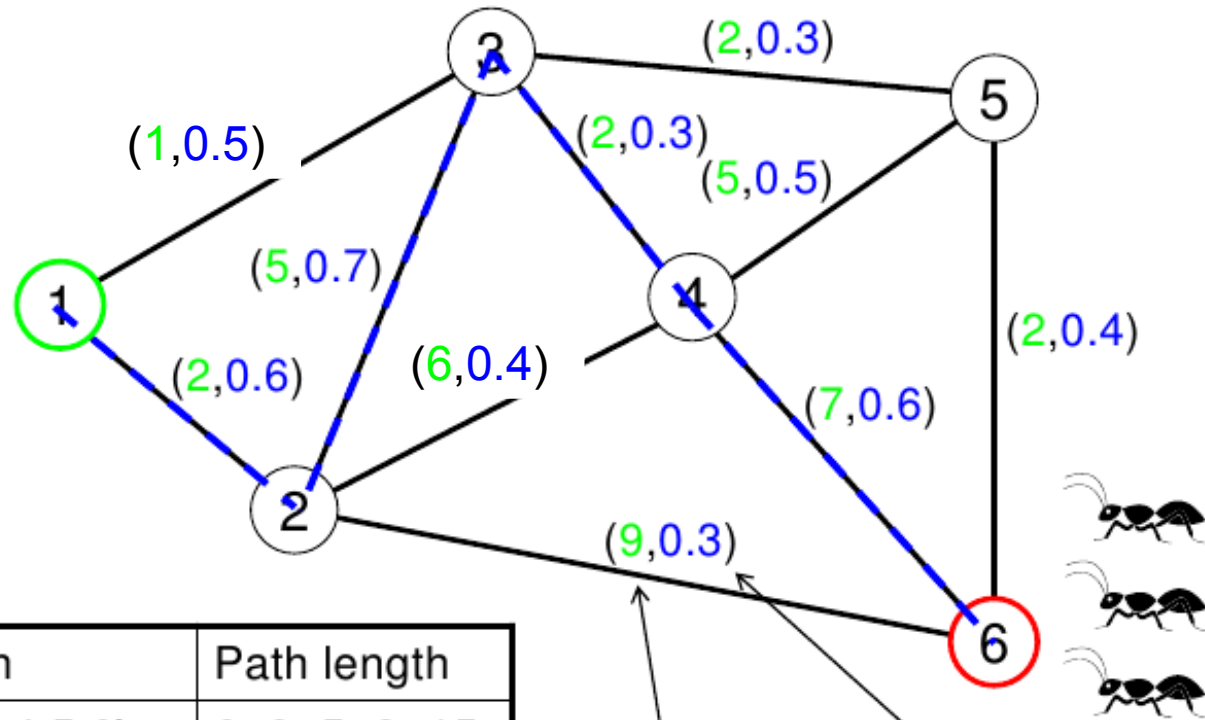
with

- ρ : the evaporation rate of the old pheromone

Simple ACO: Shortest Paths



Simple ACO: Shortest Paths



Ant	Path	Path length
1	{1,2,4,5,6}	6+2+5+2=15
2	{1,3,4,2,6}	6+2+1+9=18
3	{1,2,3,4,6}	2+5+2+7=16

Edge length

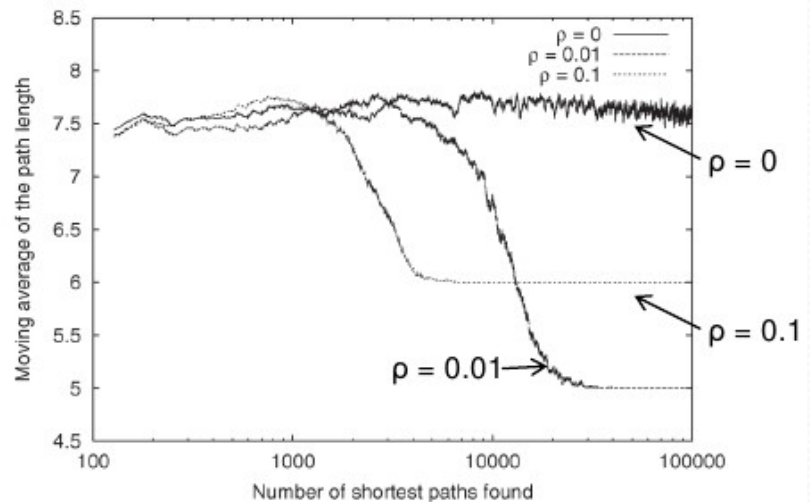
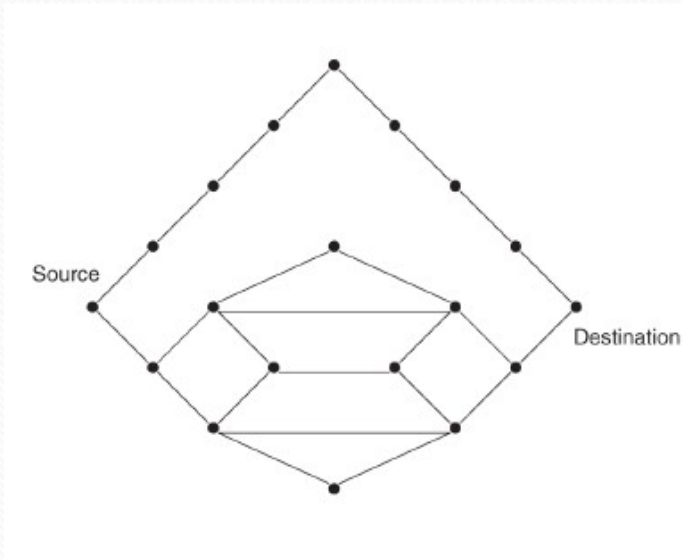
Pheromones

Simple ACO: Shortest Paths

$$Q = 1, \rho = 0.1$$

	τ_{old}	$\Delta\tau_{ij}^1$	$\Delta\tau_{ij}^2$	$\Delta\tau_{ij}^3$	$\Delta\tau_{ij}$	τ_{new}
(1,2)	0.6	1/15	0	1/16	$1/15 + 1/16 \approx 0.129$	$0.6 * 0.9 + 0.129 = 0.669$
(1,3)	0.5	0	1/18	0	$1/18 \approx 0.055$	$0.5 * 0.9 + 0.055 = 0.505$
(2,3)	0.7	0	0	1/16	$1/16 \approx 0.063$	$0.7 * 0.9 + 0.063 = 0.693$
(2,4)	0.4	1/15	1/18	0	$1/15 + 1/18 \approx 0.122$	$0.4 * 0.9 + 0.122 = 0.482$
(2,6)	0.3	0	1/18	0	$1/18 \approx 0.055$	$0.3 * 0.9 + 0.055 = 0.325$
(3,4)	0.3	0	1/18	1/16	$1/18 + 1/16 \approx 0.118$	$0.3 * 0.9 + 0.118 = 0.388$
(3,5)	0.3	0	0	0	0	$0.3 * 0.9 + 0 = 0.27$
(4,5)	0.5	1/15	0	0	$1/15 \approx 0.067$	$0.5 * 0.9 + 0.067 = 0.517$
(4,6)	0.6	0	0	1/16	$1/16 \approx 0.063$	$0.6 * 0.9 + 0.063 = 0.603$
(5,6)	0.4	1/15	0	0	$1/15 \approx 0.067$	$0.4 * 0.9 + 0.067 = 0.427$

Simple ACO: Shortest Paths



- Low ρ \rightarrow low evaporation \rightarrow slow convergence, “old” paths continue to be traversed instead of searching new ones
- High ρ \rightarrow high evaporation \rightarrow very fast convergence, but due to limited memory no drive to explore variations of a good path

Ant Systems for the Traveling Salesman Problem

- The first ACO algorithm proposed by Dorigo et al. in 1991

```
procedure Ant System for TSP
  Pheromone Initialization
  while (not terminate) do
    for i = 1 to k do
      Tour Construction
    end
    Update Pheromones
  end
end
```

Ants for TSP

- For an ant located at node v_i the probability p_{ij} of choosing v_j as the next node is:

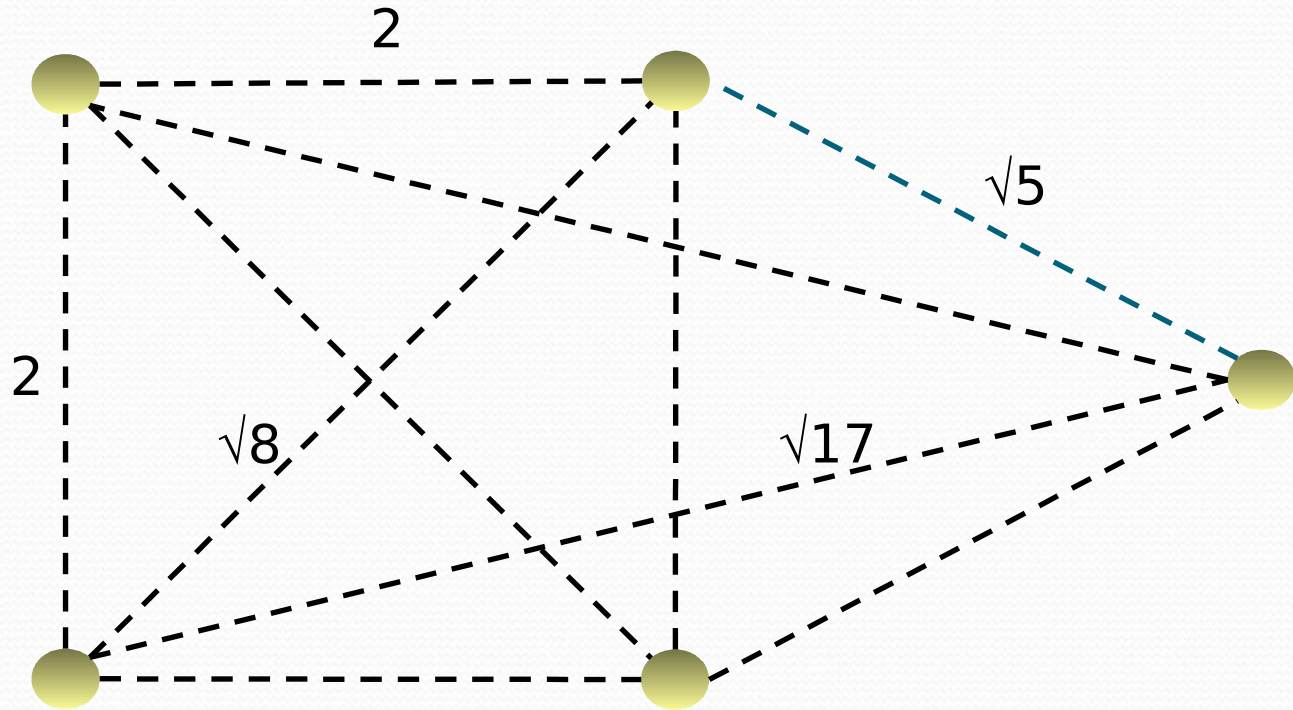
$$p_{ij}^k = \begin{cases} \frac{(\tau_{ij})^\alpha (\eta_{ij})^\beta}{\sum_{m \in N_i^k} (\tau_{im})^\alpha (\eta_{im})^\beta} & \text{if } j \in N_i^k \\ 0 & \text{if } j \notin N_i^k \end{cases}$$

where

- τ_{ij} is the amount of pheromones on edge $i \rightarrow j$
- N_i^k is the set of neighbors of node i not visited by ant k yet (tabu list)
- η_{ij} is the heuristic desirability of the edge (i.e. $1 / \text{distance between nodes}$)

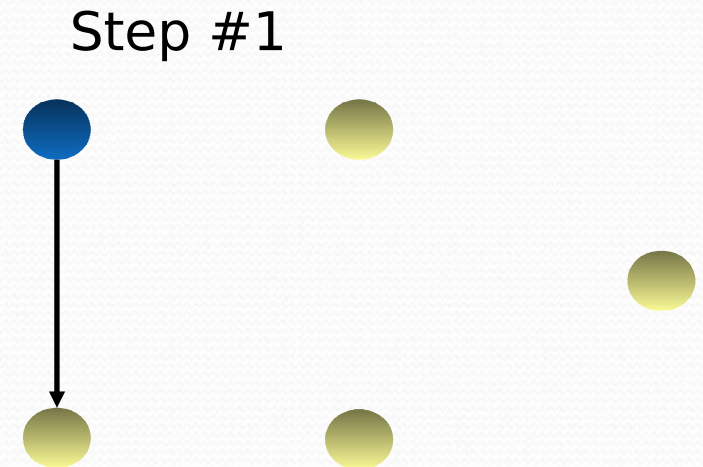
Traveling Salesman Problem

- n cities (5)
- Number of possible paths: $(n-1)! / 2$



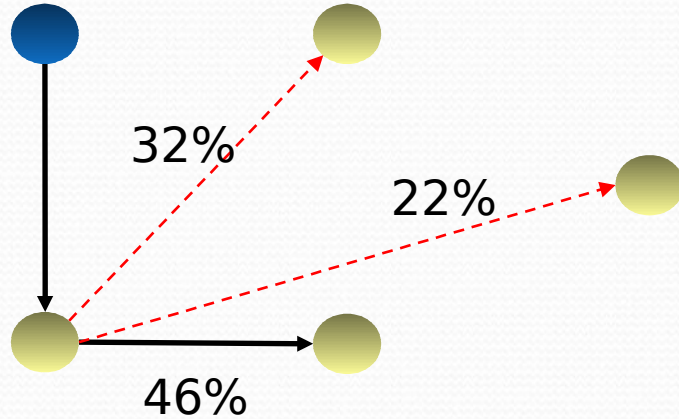
Iteration $i=1$, Ant $m=1$

- All paths have the same pheromone intensity $\tau_o=0.5$
- Pheromone trail and heuristic information have the same weight $\alpha = 1$, $\beta = 1$, $\rho=0.1$
- An ant is randomly placed
- The probability to choose is, in this case, based only on heuristic information
 - $P_{12}=31\%$
 - $P_{13}=16\%$
 - $P_{14}=22\%$
 - $P_{15}=31\%$
- Ant $m = 1$ chooses node 5

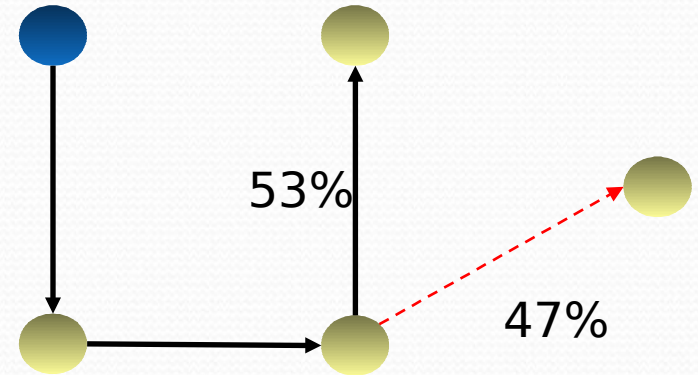


Iteration $i=1$, Ant $m=1$

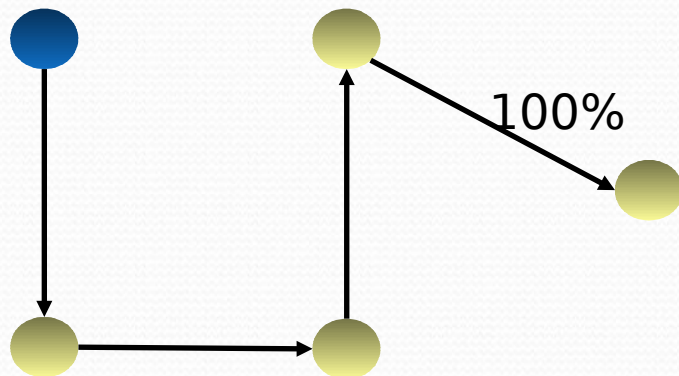
Step #2



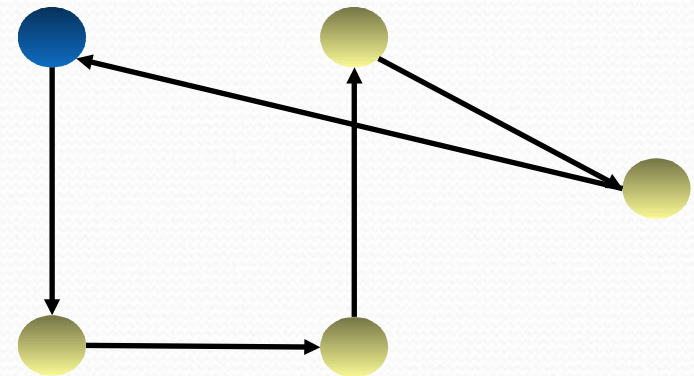
Step #3



Step #4



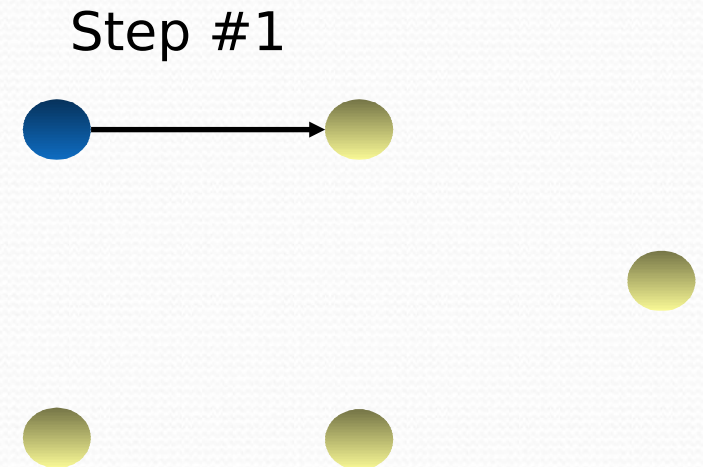
Step #5



$$f_1 = 2 + 2 + 2 + \sqrt{5} + \sqrt{17} = 12.36$$

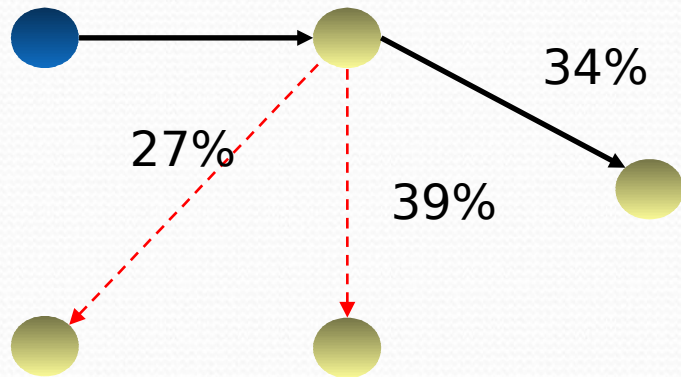
Iteration $i=1$, Ant $m=2$

- All paths have the same pheromone intensity $\tau_o=0.5$
- Pheromone trail and heuristic information have the same weight $\alpha = 1, \beta = 1, \rho=0.1$
- An ant is randomly placed
- The probability to choose is, in this case, based only on heuristic information
 - $P_{12}=31\%$
 - $P_{13}=16\%$
 - $P_{14}=22\%$
 - $P_{15}=31\%$
- Ant $m = 2$ chooses node 2

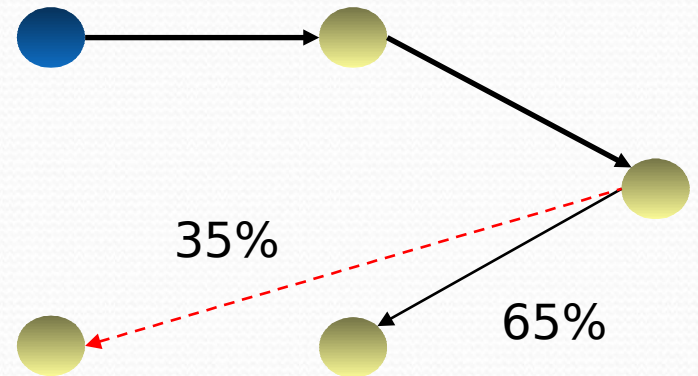


Iteration i=1, Ant m=2

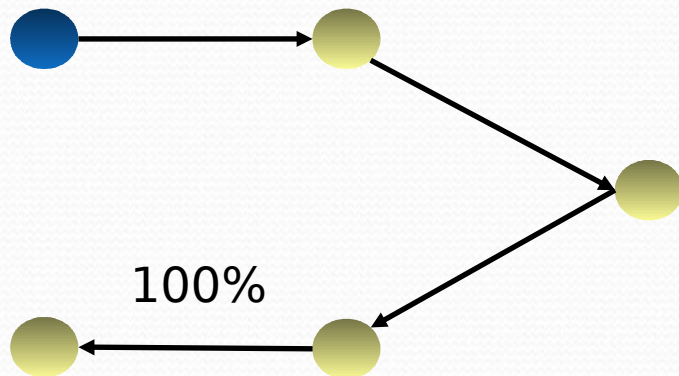
Step #2



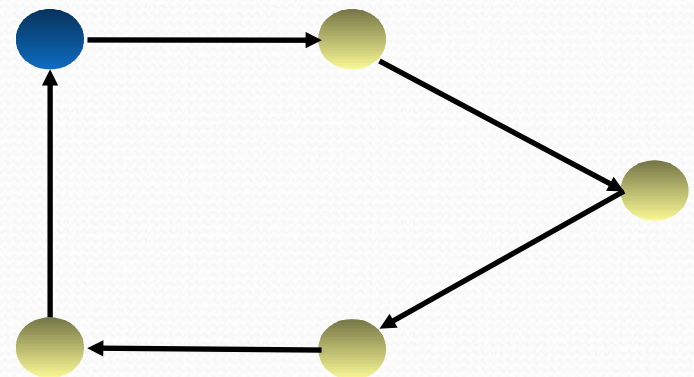
Step #3



Step #4



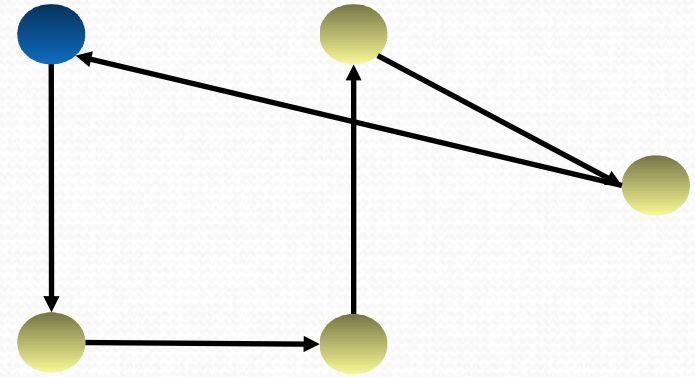
Step #5



$$f_2 = 2 + \sqrt{5} + \sqrt{5} + 2 + 2 = 10.47$$

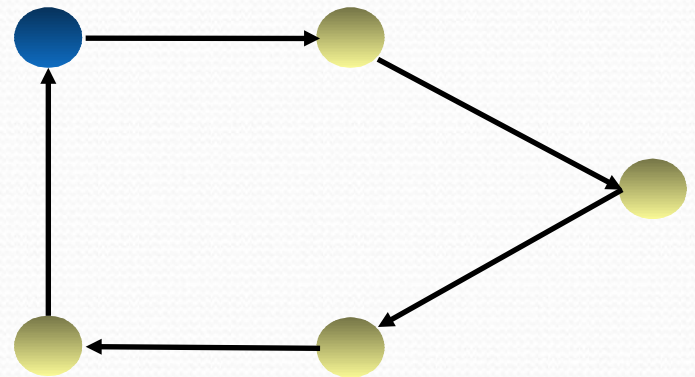
Iteration $i=1$, Pheromone Update

- The final solution of ant $m=1$ is $D=12.36$. The reinforcement produced by this ant $m=1$ is $0,081$.



$$Q = 1, \mu = Q/D$$

- The final solution of ant $m=2$ is $D=10,47$. The reinforcement produced by ant $m=2$ is $0,095$!



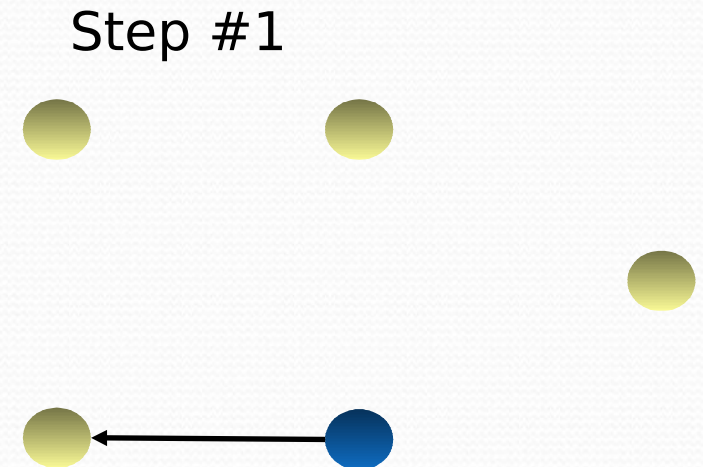
Updating Pheromone Matrix

- Update the pheromones on all edges by:

$$\tau^{(l+1)} = \begin{bmatrix} 0.5 & 0.5 & 0.5 & 0.5 & 0.5 \\ 0.5 & 0.5 & 0.5 & 0.5 & 0.5 \\ 0.5 & 0.5 & 0.5 & 0.5 & 0.5 \\ 0.5 & 0.5 & 0.5 & 0.5 & 0.5 \\ 0.5 & 0.5 & 0.5 & 0.5 & 0.5 \end{bmatrix} \times (1 - \rho) + \begin{bmatrix} 0 & 0 & 0 & 0 & 0.08 \\ 0 & 0 & 0.08 & 0 & 0 \\ 0.08 & 0 & 0 & 0 & 0 \\ 0 & 0.08 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0.08 & 0 \end{bmatrix} + \begin{bmatrix} 0 & 0.095 & 0 & 0 & 0 \\ 0 & 0 & 0.095 & 0 & 0 \\ 0 & 0 & 0 & 0.095 & 0 \\ 0 & 0 & 0 & 0 & 0.095 \\ 0.095 & 0 & 0 & 0 & 0 \end{bmatrix}$$

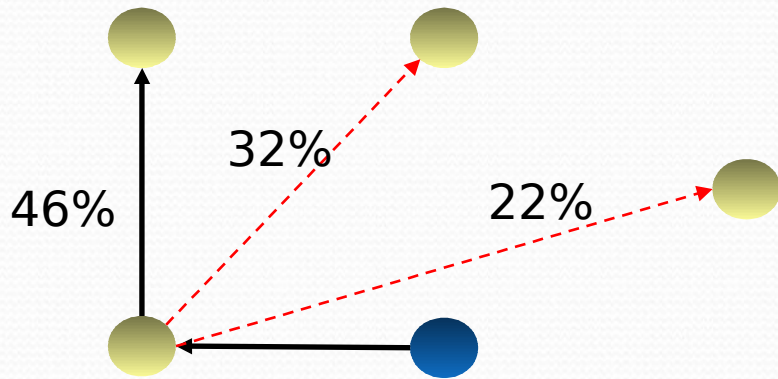
Iteration $i=2$, Ant $m=1$

- The pheromone trails have different intensities
- Pheromone trail and heuristic information have the same weight $\alpha = 1$, $\beta = 1$, $\rho=0.1$
- An ant is randomly placed
- The probability to choose is
 - $P_{41}=19\%$
 - $P_{42}=26\%$
 - $P_{43}=23\%$
 - $P_{45}=32\%$
- Ant $m = 1$ chooses node 5

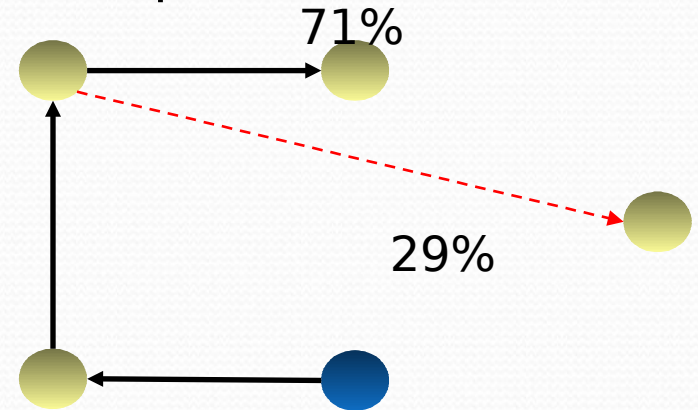


Iteration $i=2$, Ant $m=1$

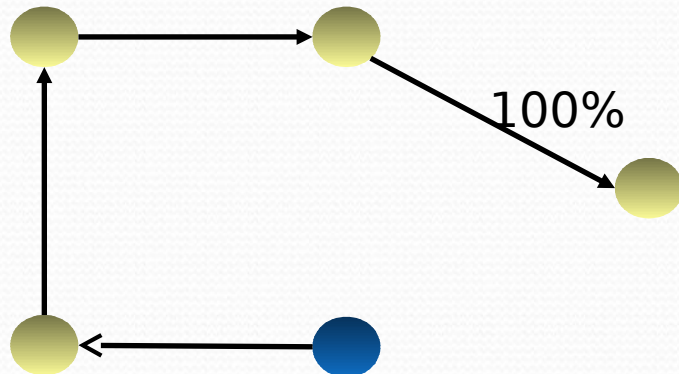
Step #2



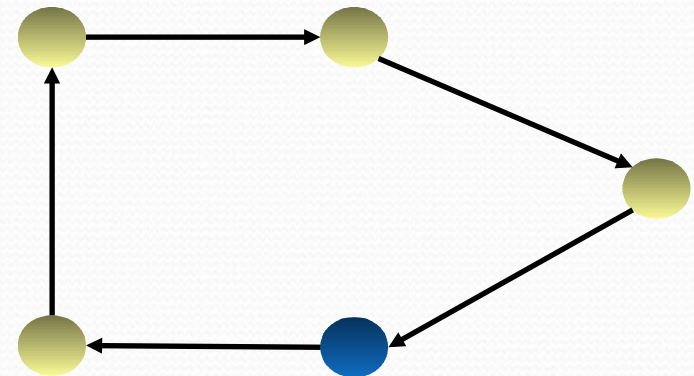
Step #3



Step #4



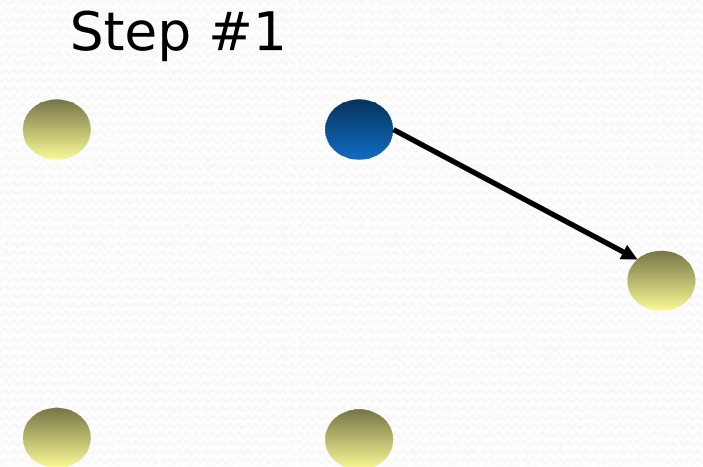
Step #5



$$f_1 = 2 + 2 + 2 + \sqrt{5} + \sqrt{5} = 10.47$$

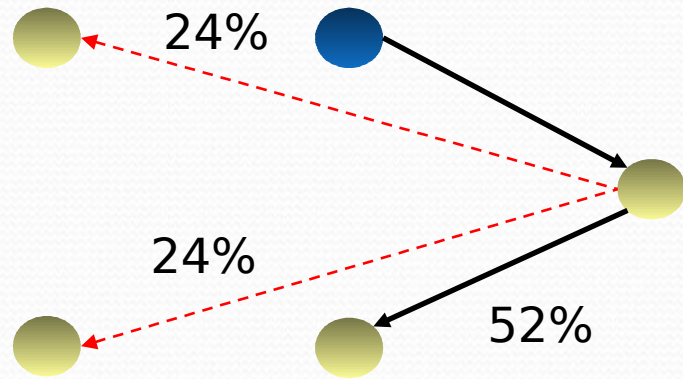
Iteration $i=2$, Ant $m=2$

- The pheromone trails have different intensities
- Pheromone trail and heuristic information have the same weight $\alpha = 1$, $\beta = 1$, $\rho=0.1$
- An ant is randomly placed
- The probability to choose is
 - $P_{21}=26\%$
 - $P_{23}=29\%$
 - $P_{24}=26\%$
 - $P_{25}=19\%$
- Ant $m = 2$ chooses node 3

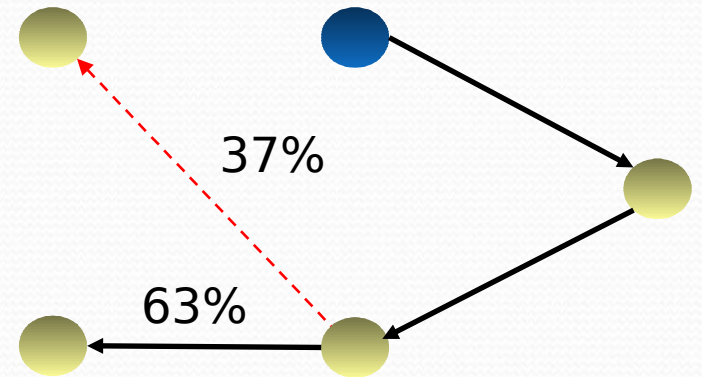


Iteration $i=2$, Ant $m=2$

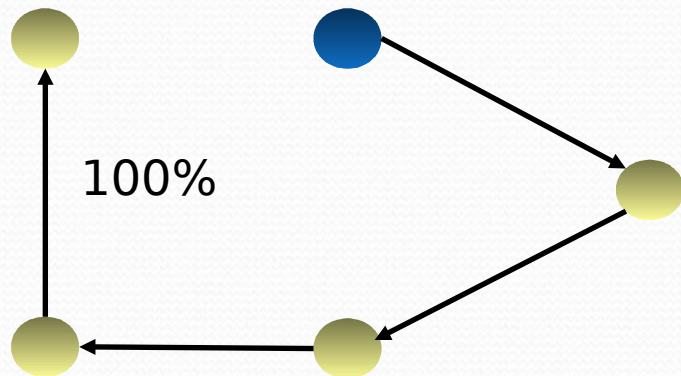
Step #2



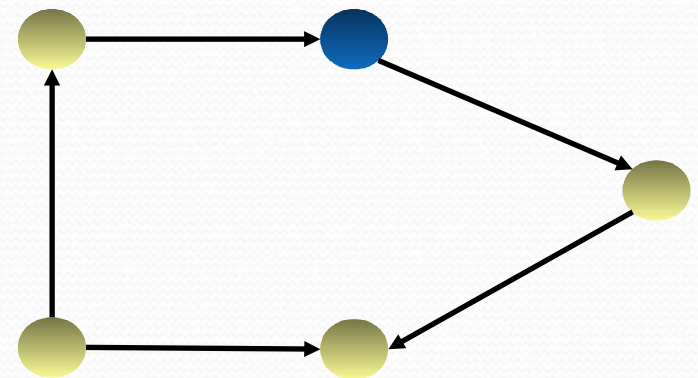
Step #3



Step #4



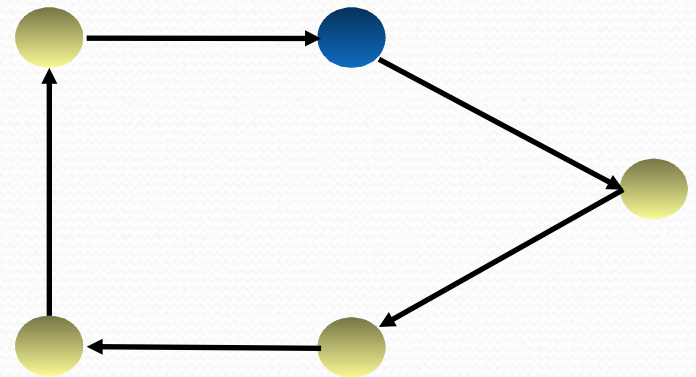
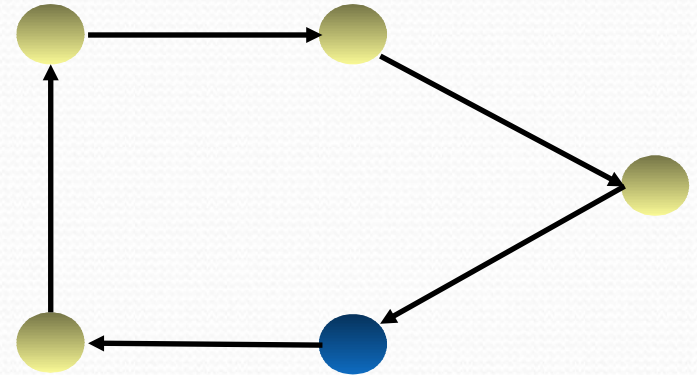
Step #5



$$f_2 = \sqrt{5+} \sqrt{5+} 2+ 2+ 2 = 10.47$$

Iteration $i=2$, Pheromone Update

- The final solution of ant $m=1$ and $m=2$ is $D=10,47$. The reinforcement produced by each ant is $0,095!$



Updating Pheromone Matrix

- Considering the pheromone dropped by every ant

$$\tau^{(l+1)} = \begin{bmatrix} 0.45 & 0.55 & 0.45 & 0.45 & 0.53 \\ 0.45 & 0.45 & 0.63 & 0.45 & 0.45 \\ 0.53 & 0.45 & 0.45 & 0.55 & 0.45 \\ 0.45 & 0.53 & 0.45 & 0.45 & 0.55 \\ 0.55 & 0.45 & 0.45 & 0.53 & 0.45 \end{bmatrix} \times (1-\rho) + \begin{bmatrix} 0 & 0.095 & 0 & 0 & 0 \\ 0 & 0 & 0.095 & 0 & 0 \\ 0 & 0 & 0 & 0.095 & 0 \\ 0 & 0 & 0 & 0 & 0.095 \\ 0.095 & 0 & 0 & 0 & 0 \end{bmatrix} + \begin{bmatrix} 0 & 0.095 & 0 & 0 & 0 \\ 0 & 0 & 0.095 & 0 & 0 \\ 0 & 0 & 0 & 0.095 & 0 \\ 0 & 0 & 0 & 0 & 0.095 \\ 0.095 & 0 & 0 & 0 & 0 \end{bmatrix}$$

ACO General Framework

Initialize pheromones

while termination conditions not met **do**

Construct ant solutions based on the pheromones

Update pheromones

Perform daemon actions (optional)

end while

Additional
local search
to improve
solutions often
necessary

Elitist Ant Systems

- Modification of the ant system in which the best-so-far solution is reinforced additionally:

$$\tau_{ij} = (1 - \rho)\tau_{ij} + \sum_{k=0}^m \Delta\tau_{ij}^k + e \cdot \Delta\tau_{ij}^{bs}$$

With

- *bs*: the best solution found till now
- *e*: parameter for the contribution of the best solution

Example:

Bankruptcy Prediction

- Bankruptcy prediction is a classification problem:
find a classification rule that will separate firms that will go bankrupt from those that will not
- The set of attributes is usually a set of financial variables
- Most successful breakthrough in BP by Altman, 1968

Bankruptcy Prediction

- Altman selected in first instance 5 variables out of a list of 22 financial variables.

X_1 : Working Capital / Total Assets

X_2 : Retained Earnings / Total Assets

X_3 : EBIT / Total Assets

X_4 : MV of Equity / BV of Debt

X_5 : Sales / Total Assets

$$Z = .012X_1 + .014X_2 + .033X_3 + .006X_4 + .999X_5$$



Altman's Data

- The dataset used by Altman consisted of 66 companies, 33 bankrupt (B) and 33 non bankrupt (NB)

Bankrupt

- Asset size between 0.6 mil. and 25.9 mil.
- Filed for bankruptcy between 1946 – 1965
- Using data for the 5 variables from 1 year before filing for bankruptcy

Non-Bankrupt

- Asset size between 1 mil. and 25 mil.
- Still in existence in 1966

Formalisation as Discrete Optimisation Problem

- For each variable (attribute) in the analysis, we generate cutpoints to discretise the data
- All possible cutpoints for a variable X_i are obtained by dividing the interval $[\min(i), \max(i)]$ into a fixed number of smaller intervals
 - For each variable i we have cutpoints j , θ_{ij}
- For each variable i we have to choose one θ_{ij}

Formalisation as Discrete Optimisation Problem

- Evaluation of a choice of cutpoints:
 - we predict bankruptcy for a firm k with attributes

$$\xi_1^k, \dots, \xi_n^k$$

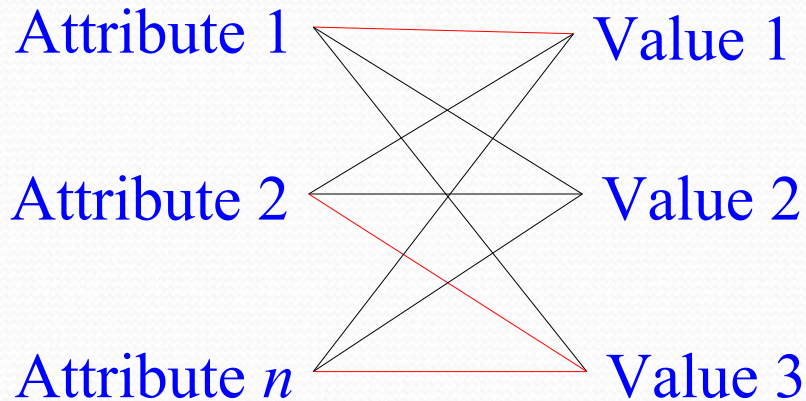
$$\text{if } \xi_1^k \leq \theta_{1c(1)} \wedge \dots \wedge \xi_n^k \leq \theta_{nc(n)}$$

where $c(i)$ is the cutpoint chosen for attribute i

- quality of a solution: the error of this choice of cutpoints on the training data

Ant Optimisation Representation

- We can see assignments as a choice of edges in a bipartite graph → update pheromones for each edge



Ant Optimisation

Representation

- Pheromone update for **ant** k

$$\Delta_{ij}^k = \begin{cases} A & \text{if } c(i) = j \\ 0 & \text{otherwise} \end{cases}$$

where A is the the number of correctly predicted training examples

- Ants search for solutions by choosing the cutpoint for each variable in a fixed order

Ant Optimisation Representation

- Define a (heuristic) distance to each cutpoint for the next variable:

$$\eta_{ij} = \text{accuracy when only using attributes } i-1 \text{ and } i \text{ with cut points } c(i-1) \text{ and } j$$

→ large value is more promising

- Otherwise equal to ant systems for the TSP

Experiments

- We employ 2 datasets:
 - The Altman dataset
 - A custom dataset consisting of:
 - 110 firms (55 B and 55 NB)
 - The firms filed for bankruptcy between 1998 and 2004
 - Asset size lower than 1 billion when filing for bankruptcy
 - Using data 2 years prior to bankruptcy
 - The NB set contains firms still 'alive' in 2005

Experiments

- The parameters used:
 - $\alpha = 1$
 - $\beta = 1$
 - $\rho = 0.5$
 - 30 ants on the Altman dataset, 40 on the second
 - Different experiments have been performed, using the whole dataset or dividing the latter in a training and test subset.
- Comparison with *multiple discriminant analysis*, used by Altman and the most popular method

Results

type 1 err.	type 2 err.	hitrate	st. dev.	min.	max.
2.8 (8.5%)	1.6 (4.8%)	93.3%	1.06	92.4%	95.5%
2.0 (6.1%)	1.0 (3.0%)	95.5%	N/A	N/A	N/A

TABLE I

RESULTS OBTAINED WITH THE COMPLETE DATA SET 1, THE FIRST ROW USING THE AA, THE SECOND ROW USING MDA.

type 1 err.	type 2 err.	hitrate	st. dev.	min.	max.
11.4 (20.7%)	3.7 (6.7%)	86.3%	0.9	85.5%	87.3%
12.0 (21.8%)	10.0 (18.2%)	80.0%	N/A	N/A	N/A

TABLE III

RESULTS OBTAINED WITH THE COMPLETE DATA SET 2, THE FIRST ROW USING THE AA, THE SECOND ROW USING MDA.

Results

	TRAINING SET	TEST SET
AA	95.2%	90.8%
MDA	97.6%	95.8%

TABLE II

RESULTS WITH DATA SET 1, USING A SEPARATE TRAINING AND TEST SET.

	TRAINING SET	TEST SET
AA	87.1%	81.0%
MDA	77.1%	70.0%

TABLE IV

RESULTS WITH DATA SET 2, USING A SEPARATE TRAINING AND TEST SET.

Results

train 1	test 1	train 2	test 2	train 3	test 3
86.9%	79.5%	88.3%	76.5%	85.7%	76.0%
77.1%	72.7%	81.4%	80.0%	92.9%	72.5%

TABLE V

RESULTS OF THE EXTRA EXPERIMENTS WITH DATA SET 2, THE FIRST ROW WITH THE AA, THE SECOND ROW WITH MDA.