

# MODELING MOLECULAR PATHWAYS BASED ON GENE EXPRESSION AND SOCIAL NETWORK ANALYSES: AN EXAMPLE FROM *ARABIDOPSIS THALIANA*

Di Salle P.\*, Colantuono C.\*, Gallo C.\*\* , Traini A.\* , Frusciante L.\* . Chiusano M. L.\*

\*Dept. of Soil, Plant, Environmental and Animal Production Sciences

\*\* University of Foggia, Dept. of Clinical and Experimental Medicine

Contact: [pasquale.disalle@libero.it](mailto:pasquale.disalle@libero.it)  
PI: [chiusano@unina.it](mailto:chiusano@unina.it)

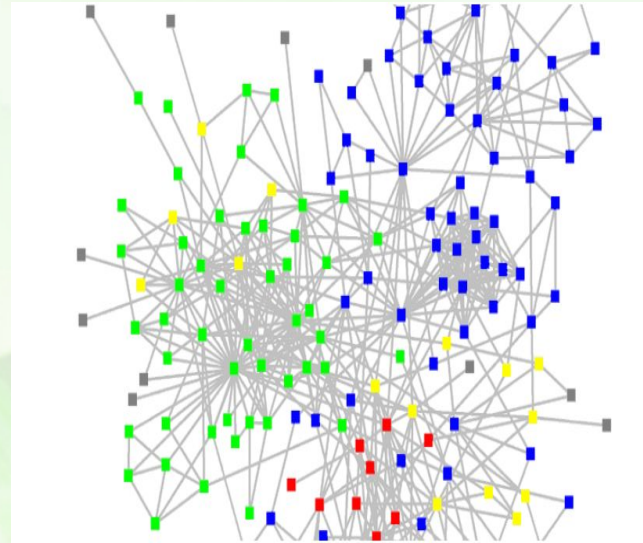
# Introduction

## ***Arabidopsis thaliana...***

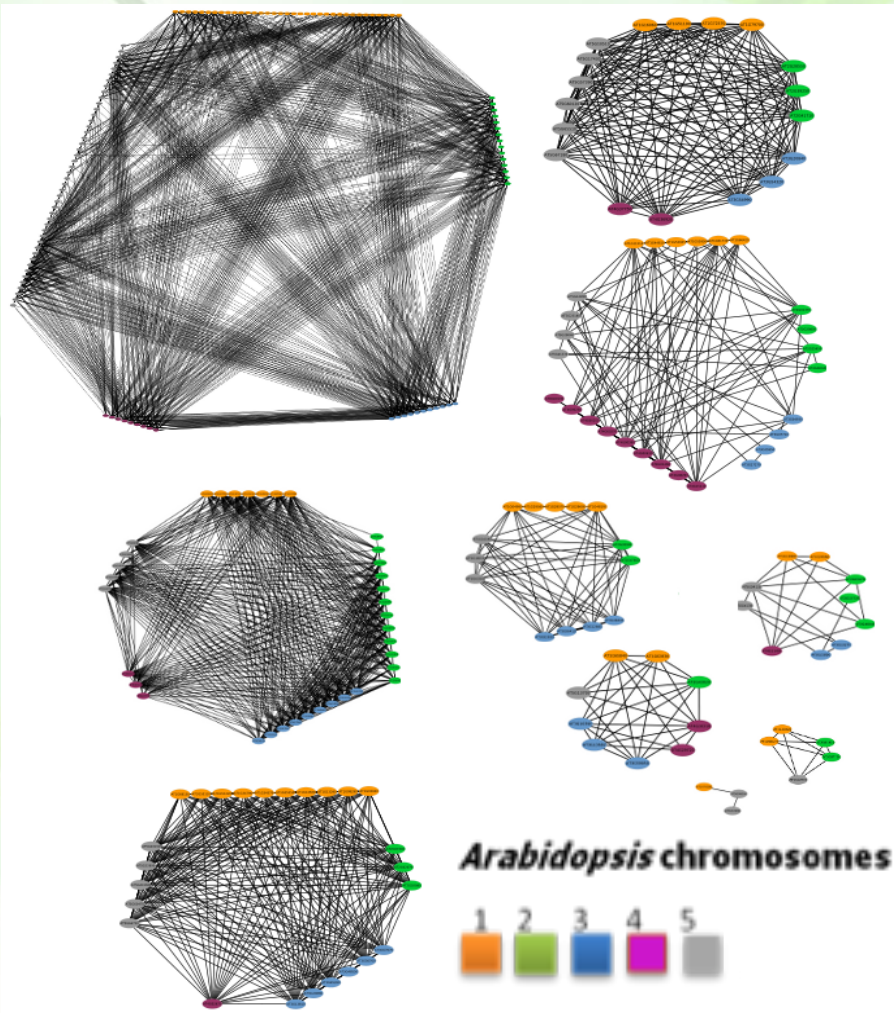
- *Reference genome model for plant biology*
- *Several rounds of genome duplication*
- *High presence of paralogs → hard function assignment*

## ***Aims:***

*Associate global coexpression analyses with the paralog organization in Arabidopsis thaliana*



# Re-organization of *A. thaliana* genome in terms of duplicated genes



Gene number in a paralog net	Paralog Net Number	% Protein Coding Genes
2	1347	10
3-9	1370	23
10-30	216	13
31-207	83	15
208-5168	1	19
<b>21843</b>	<b>3017</b>	<b>80</b>

**80%** of coding gene has at least one paralog  
**20%** of the protein coding genome is in single copy\*

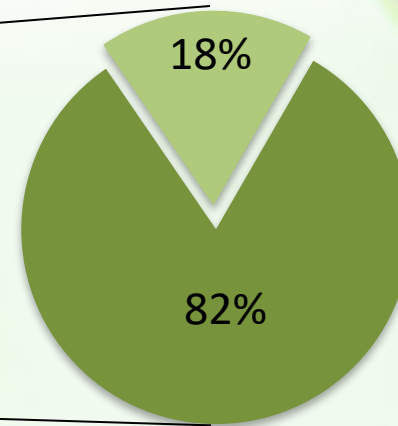
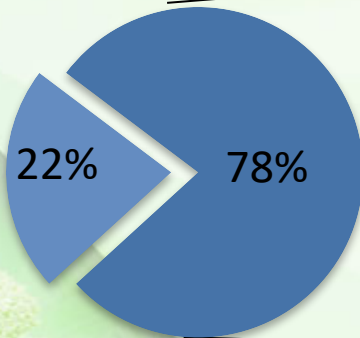
Only 7518 genes , for 380 families, are annotated in literature

\*“Exploiting a reference genome in terms of duplications network of paralogs and single copy genes in *Arabidopsis thaliana*” Vigilante A., SanGiovanni M., Frusciante L., Chiusano M. L., *In press*

# Correlation analysis on 189 microarrays from Atgen Express Project

Nr. of microarray slices	Tissue	Age	Conditions	Substrate
45	Flower	21 days	flower stage 9-15	Soil
54	Leaves	7-21 days	plant with 1-12 leaves	Soil
21	Shoots & Stems	7-21 days	shoot apex, inflorescence	Soil
24	Siliques and seeds	8 wks	siliques/seeds stage 1-10	Soil
21	Roots	7-21 days	roots	Soil
24	Seedlings and whole plant	7-23 days	seedlings and green parts	Soil

- Probes
- No Probes



- Genes with  $R > 0.7$
- Genes with  $R < 0.7$

82% of coding genes with probes makes at least one correlation with another gene

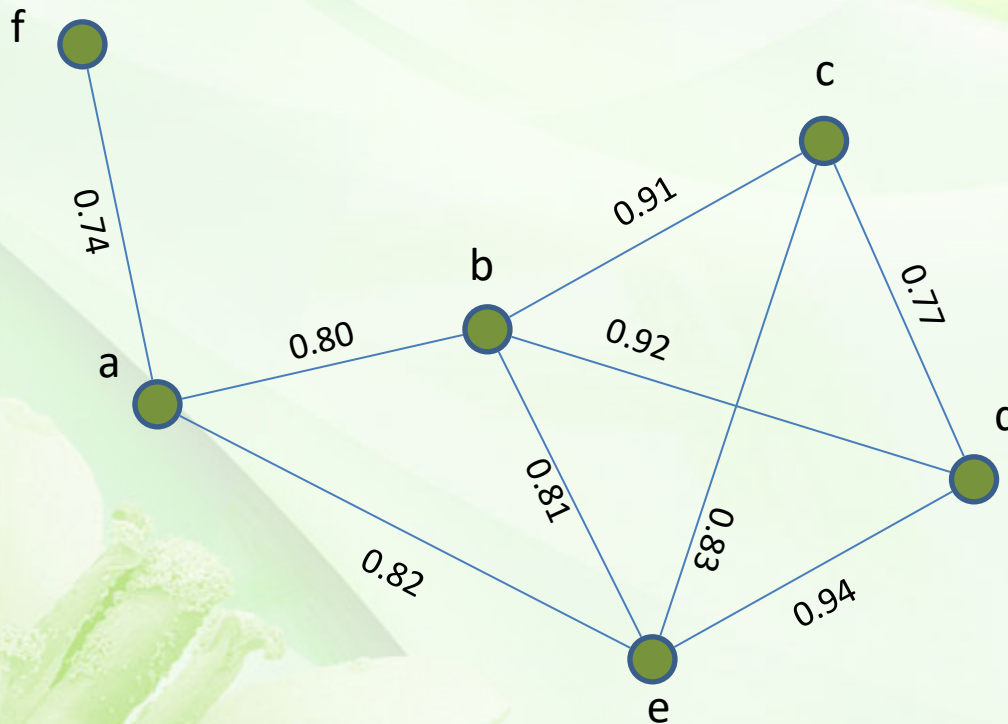
Total protein coding genes in TAIR	27169	(%)
-genes with probes in Affymetrix chip	21176	77.94
-genes without probes in Affymetrix chip	5993	22.06

Genes with probes	21176	(%)
-genes with $R > 0.7$	17386	82.20
-genes with $R < 0.7$	3790	17.80

# A social network algorithm...

ALL NODES WITH  $\vartheta_1 > 0.70$  ARE SHOWN AS A CLUSTER

$\vartheta_1$  = Pearson's R value



**Ref.** Albert, Barabasi (Review of Modern Physics, Vol.74, 2002) Statistical Mechanics of Complex Networks

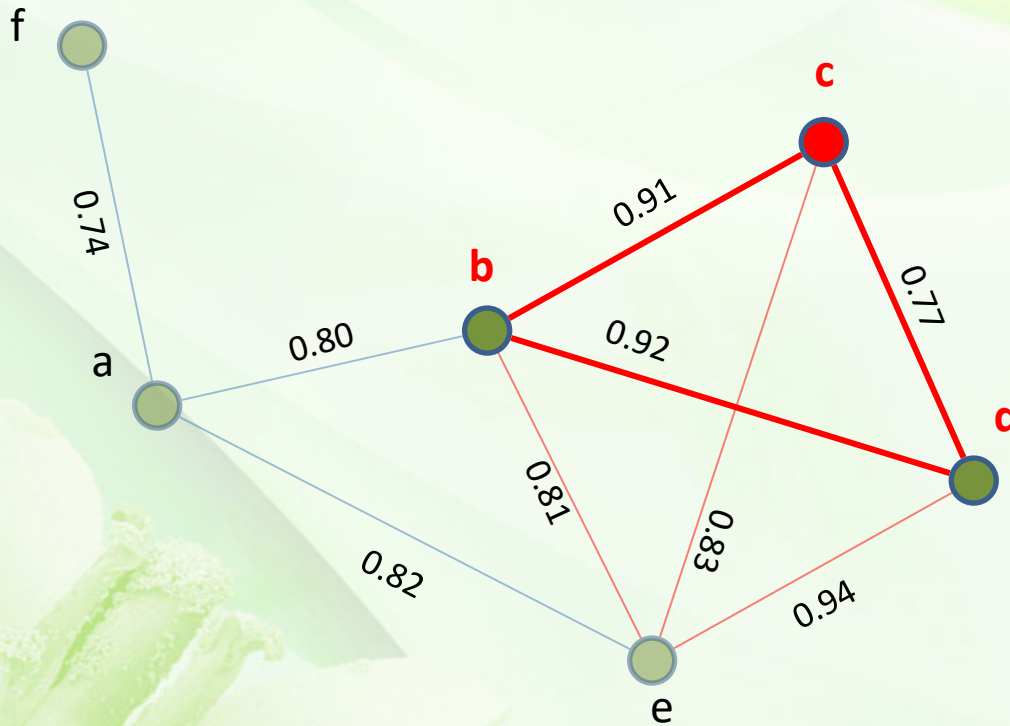
**Ref.** Barabasi, Oltvai (Nature Reviews-Genetics, Vol.5, 2004) Network Biology: Understanding the cell's functional organization

# A social network algorithm...

ALL NODES WITH  $\vartheta_1 > 0.70$  ARE SHOWN AS A CLUSTER

$\vartheta_1$  = Pearson's R value

$\vartheta_2$  = Established triangles/  
Possible triangles

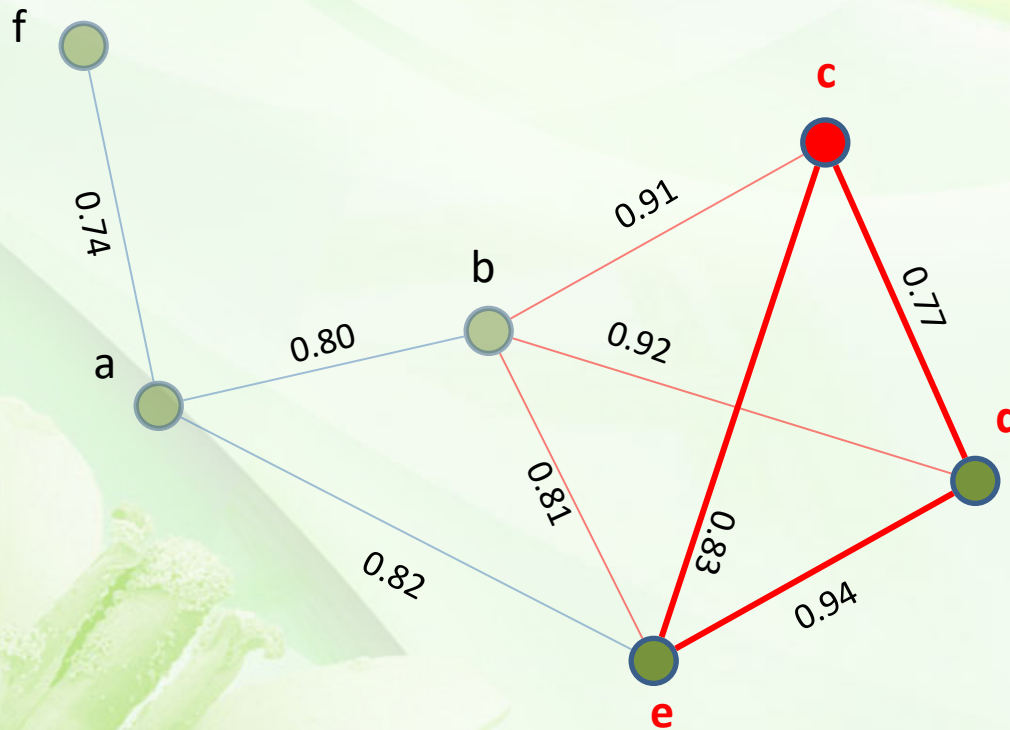


# A social network algorithm...

ALL NODES WITH  $\vartheta_1 > 0.70$  ARE SHOWN AS A CLUSTER

$\vartheta_1$  = Pearson's R value

$\vartheta_2$  = Established triangles/  
Possible triangles

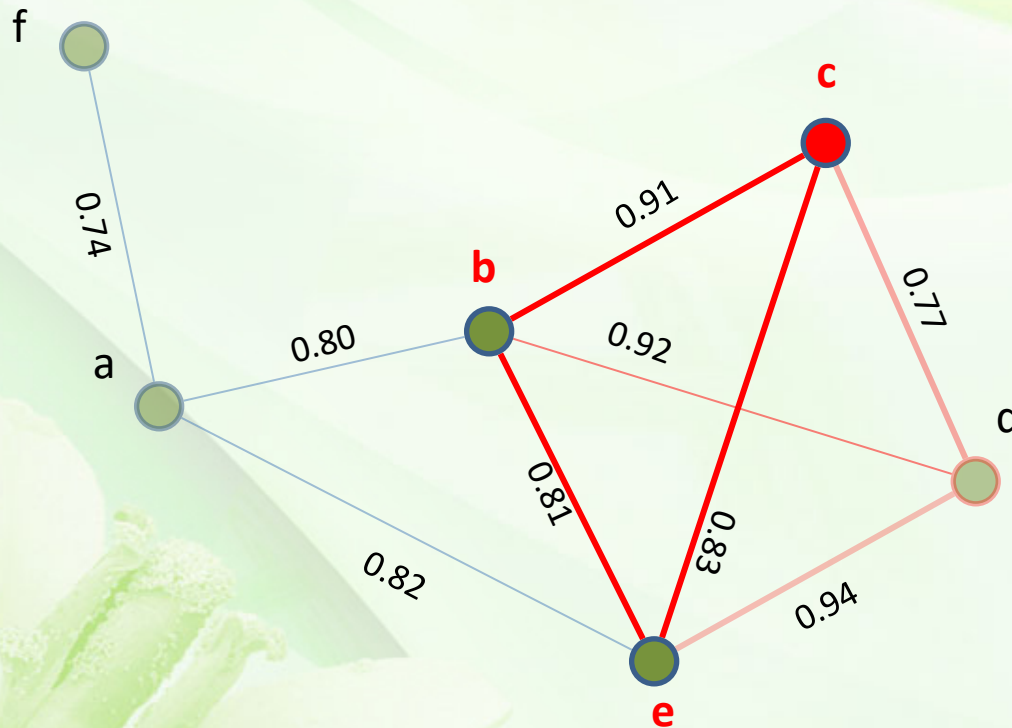


# A social network algorithm...

ALL NODES WITH  $\vartheta_1 > 0.70$  ARE SHOWN AS A CLUSTER

$\vartheta_1$  = Pearson's R value

$\vartheta_2$  = Established triangles /  
Possible triangles



$$c = 3/3 = 1$$

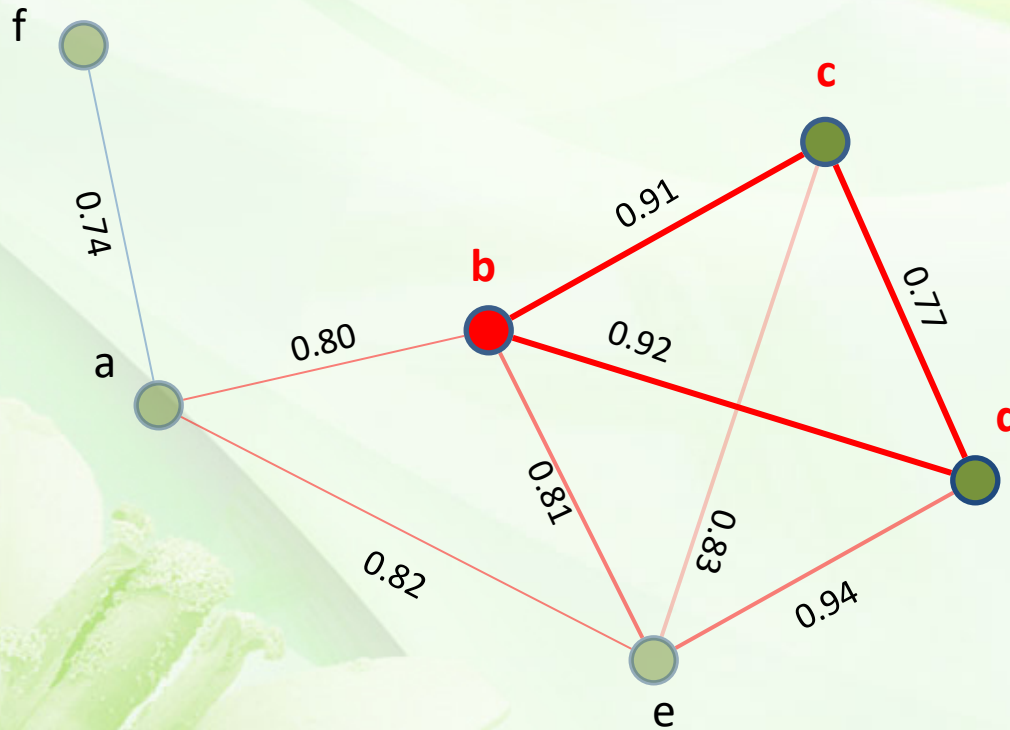


# A social network algorithm...

ALL NODES WITH  $\vartheta_1 > 0.70$  ARE SHOWN AS A CLUSTER

$\vartheta_1$  = Pearson's R value

$\vartheta_2$  = Established triangles /  
Possible triangles



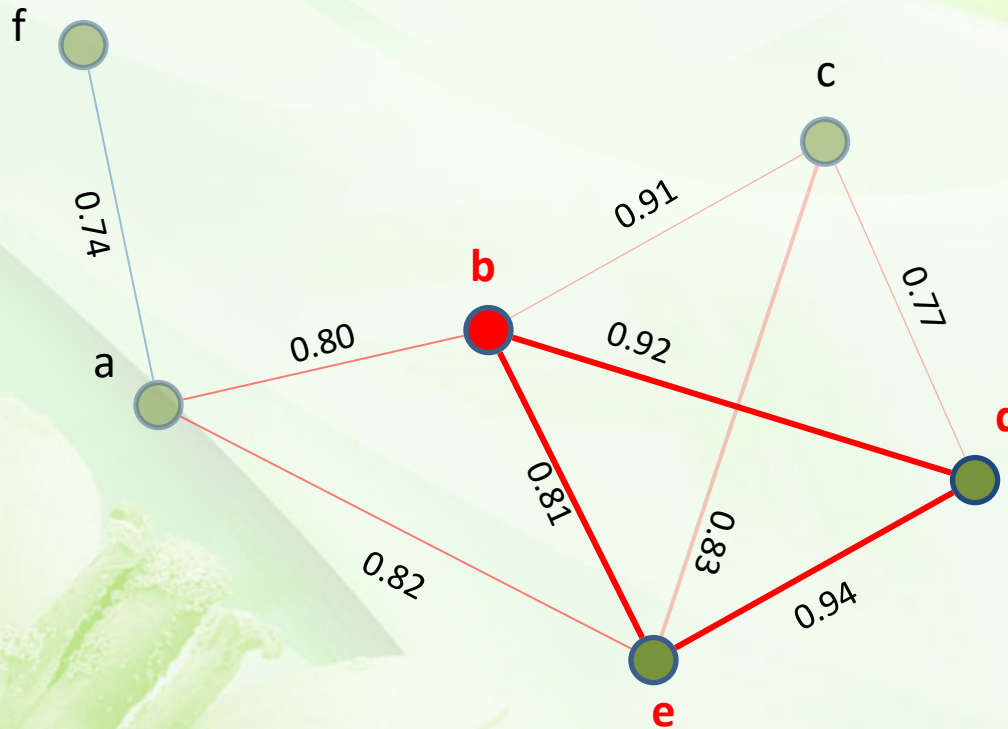
$$c = 3/3 = 1$$

# A social network algorithm...

ALL NODES WITH  $\vartheta_1 > 0.70$  ARE SHOWN AS A CLUSTER

$\vartheta_1$  = Pearson's R value

$\vartheta_2$  = Established triangles/  
Possible triangles



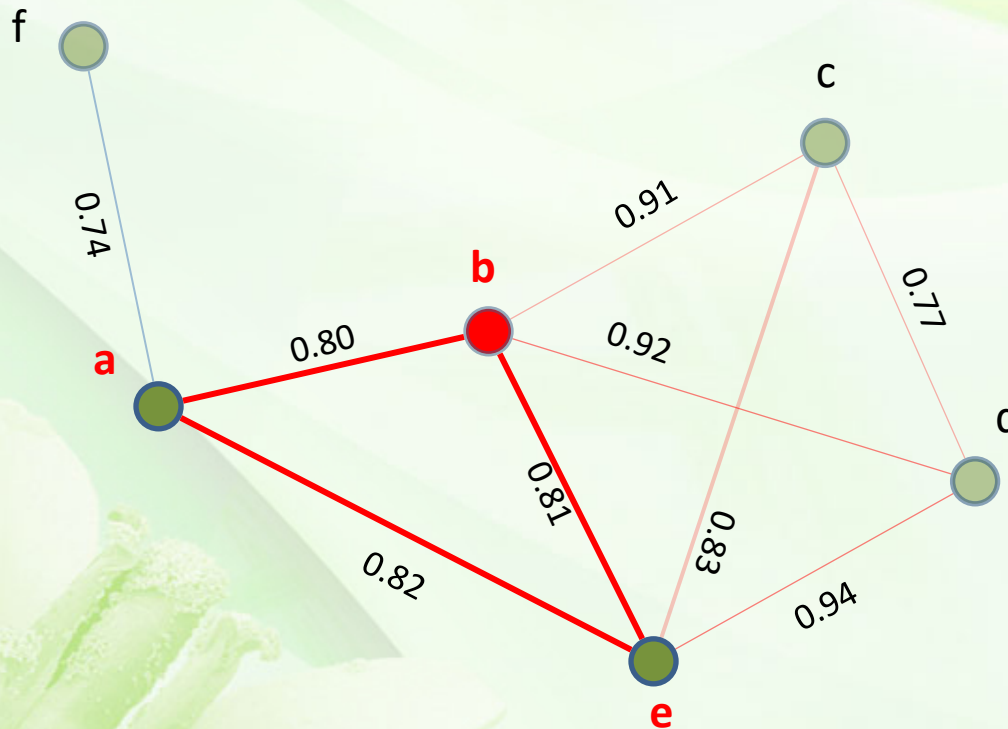
$$c = 3/3 = 1$$

# A social network algorithm...

ALL NODES WITH  $\vartheta_1 > 0.70$  ARE SHOWN AS A CLUSTER

$\vartheta_1$  = Pearson's R value

$\vartheta_2$  = Established triangles/  
Possible triangles



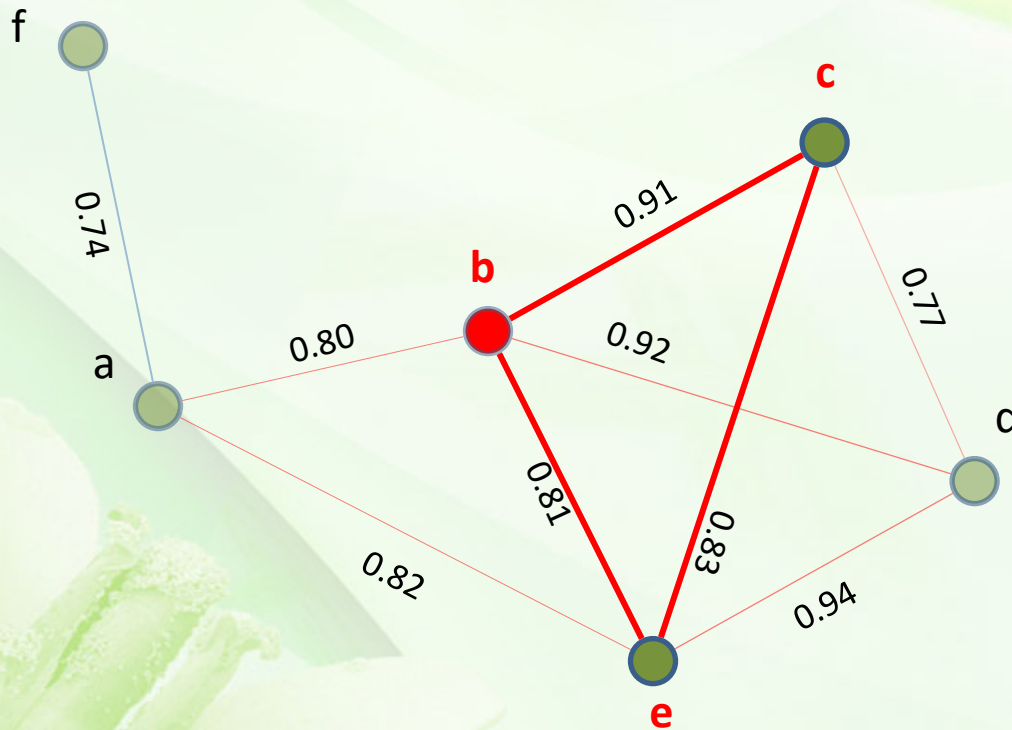
$$c = 3/3 = 1$$

# A social network algorithm...

ALL NODES WITH  $\vartheta_1 > 0.70$  ARE SHOWN AS A CLUSTER

$\vartheta_1$  = Pearson's R value

$\vartheta_2$  = Established triangles/  
Possible triangles



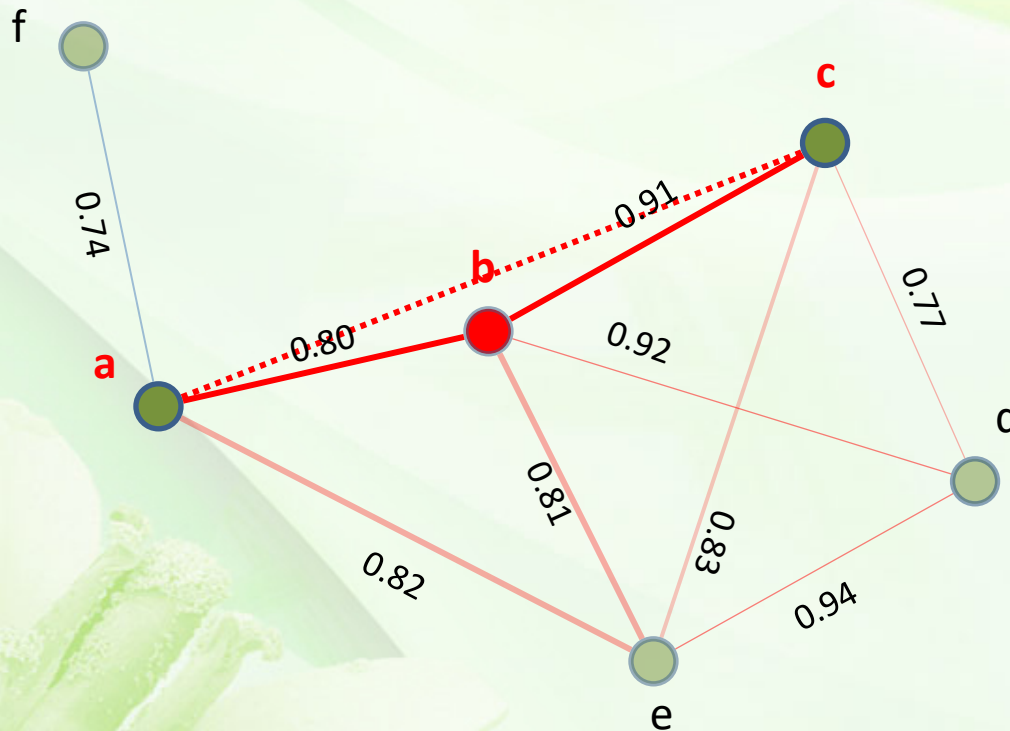
$$c = 3/3 = 1$$

# A social network algorithm...

ALL NODES WITH  $\vartheta_1 > 0.70$  ARE SHOWN AS A CLUSTER

$\vartheta_1$  = Pearson's R value

$\vartheta_2$  = Established triangles/  
Possible triangles



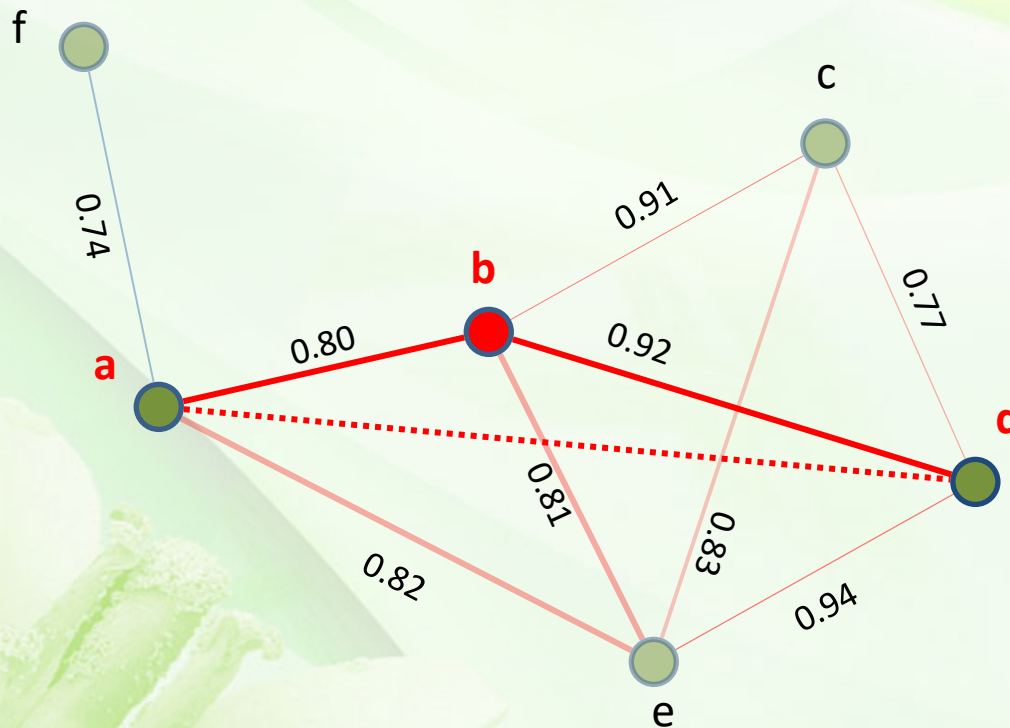
$$c = 3/3 = 1$$

# A social network algorithm...

ALL NODES WITH  $\vartheta_1 > 0.70$  ARE SHOWN AS A CLUSTER

$\vartheta_1$  = Pearson's R value

$\vartheta_2$  = Established triangles/  
Possible triangles



$$c = 3/3 = 1$$

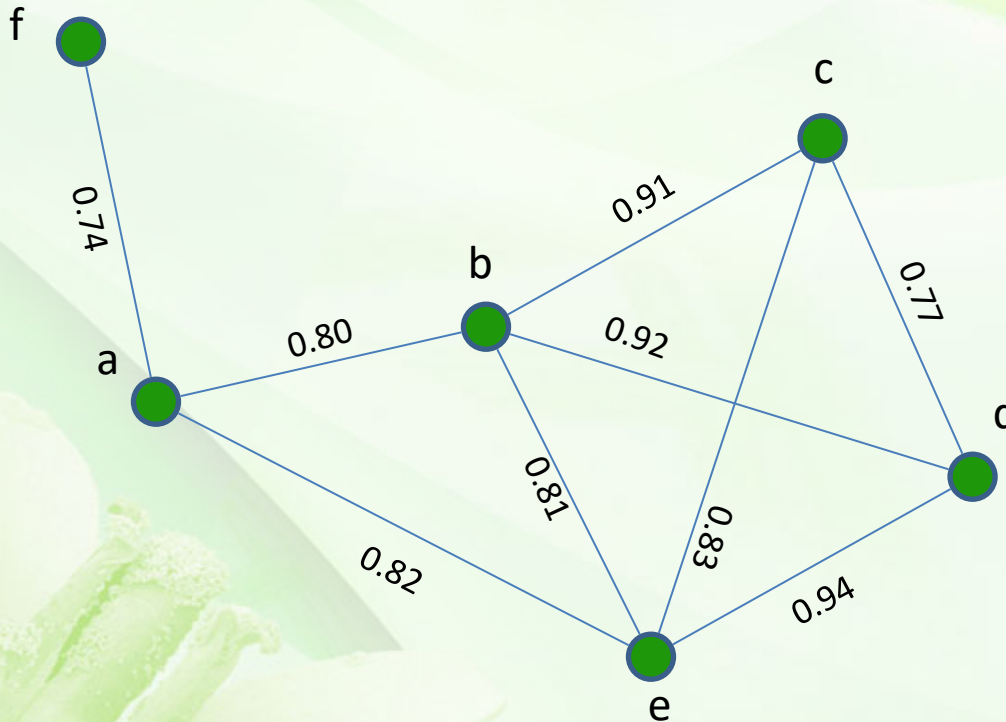
$$b = 4/6 = 0.6$$

# A social network algorithm...

ALL NODES WITH  $\vartheta_1 > 0.70$  ARE SHOWN AS A CLUSTER

$\vartheta_1$  = Pearson's R value

$\vartheta_2$  = Established triangles/  
Possible triangles



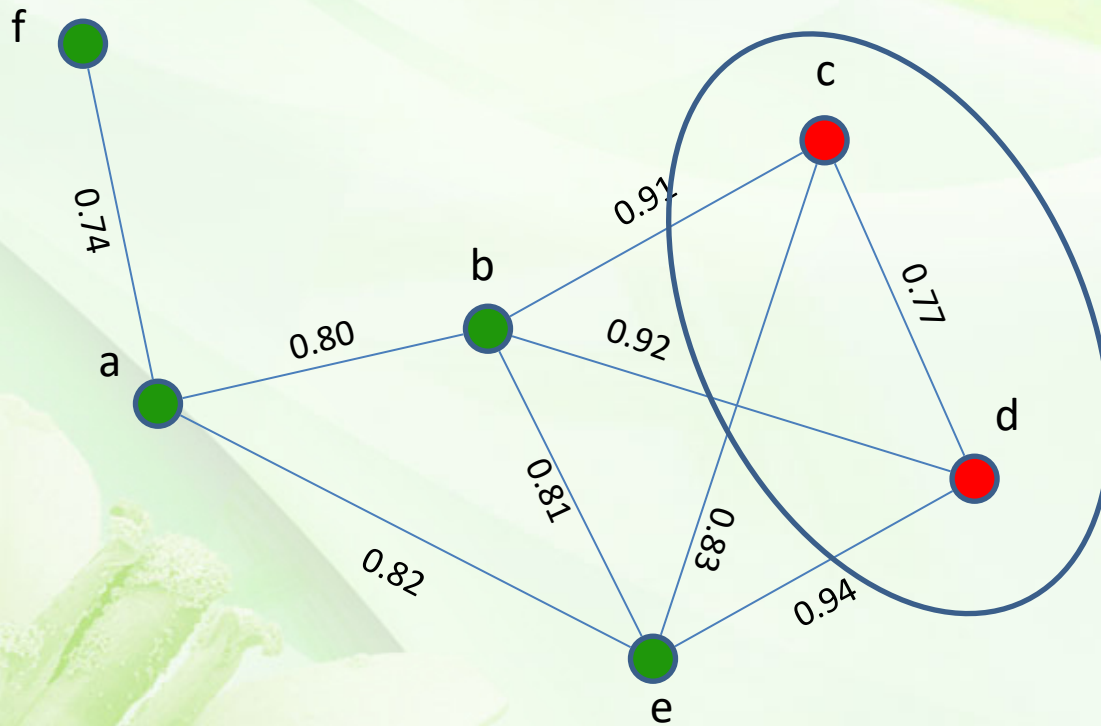
$$\begin{aligned}c &= 3/3 = 1 \\ b &= 4/6 = 0.6 \\ d &= 3/3 = 1 \\ e &= 4/6 = 0.6 \\ a &= 1/3 = 0.3 \\ f &= 0\end{aligned}$$

# A social network algorithm...

ALL NODES WITH  $\vartheta_1 > 0.70$  ARE SHOWN AS A CLUSTER

$\vartheta_1$  = Pearson's R value

$\vartheta_2$  = Established triangles /  
Possible triangles



$\vartheta_2 = 1.0$

$\vartheta_{2/gene}$	1,0
a	
b	
c	1
d	1
e	
f	

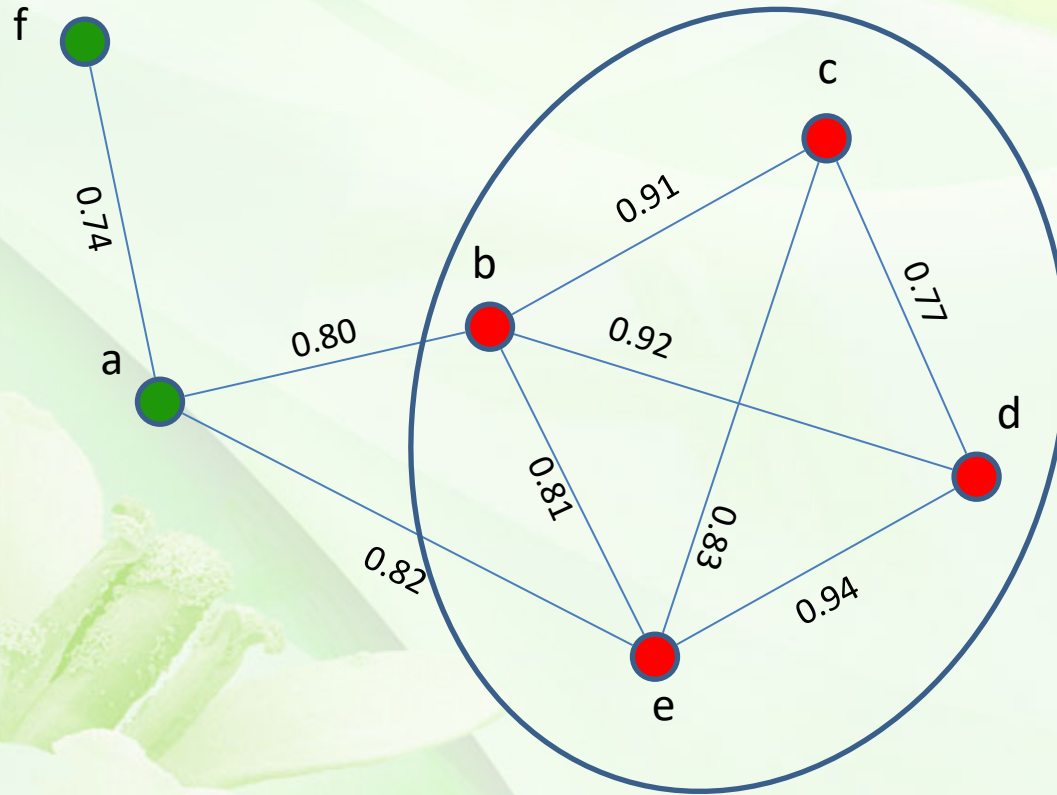


# A social network algorithm...

ALL NODES WITH  $\vartheta_1 > 0.70$  ARE SHOWN AS A CLUSTER

$\vartheta_1$  = Pearson's R value

$\vartheta_2$  = Established triangles /  
Possible triangles



$\vartheta_2 = 0.5$

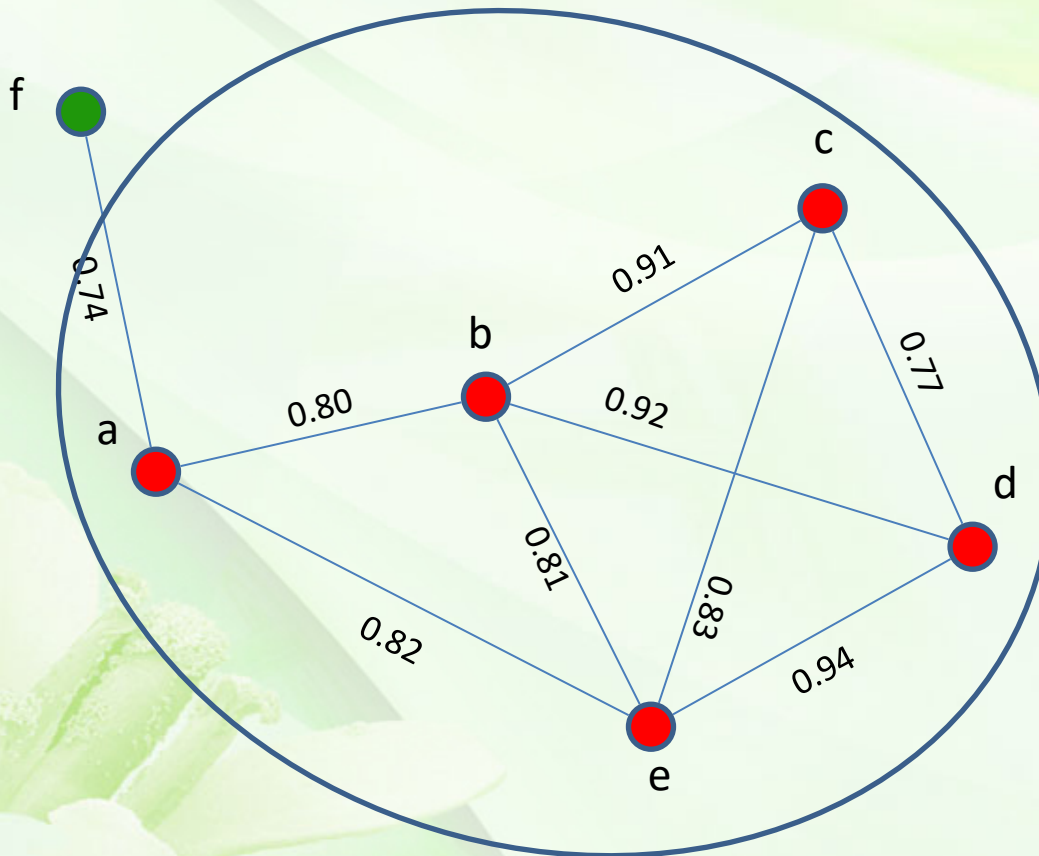
$\vartheta_{2/gene}$	0,5	1,0
a		
b	1	
c	1	1
d	1	1
e	1	
f		

# A social network algorithm...

ALL NODES WITH  $\vartheta_1 > 0.70$  ARE SHOWN AS A CLUSTER

$\vartheta_1$  = Pearson's R value

$\vartheta_2$  = Established triangles /  
Possible triangles

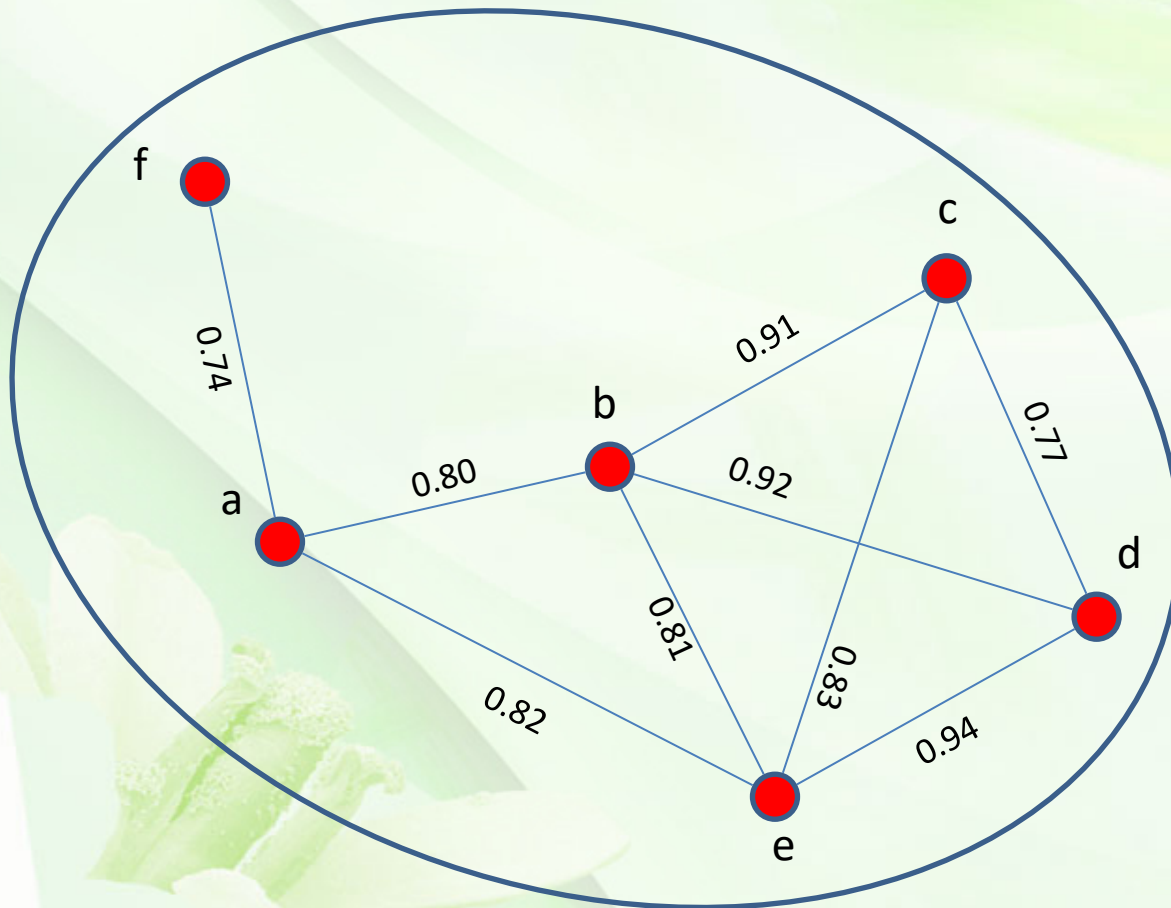


$\vartheta_2 = 0.1$

$\vartheta_{2/gene}$	0,1	0,5	1,0
a	1		
b	1	1	
c	1	1	1
d	1	1	1
e	1	1	
f			

# A social network algorithm...

ALL NODES WITH  $\vartheta_1 > 0.70$  ARE SHOWN AS A CLUSTER



$\vartheta_1$  = Pearson's R value

$\vartheta_2$  = Established triangles/  
Possible triangles

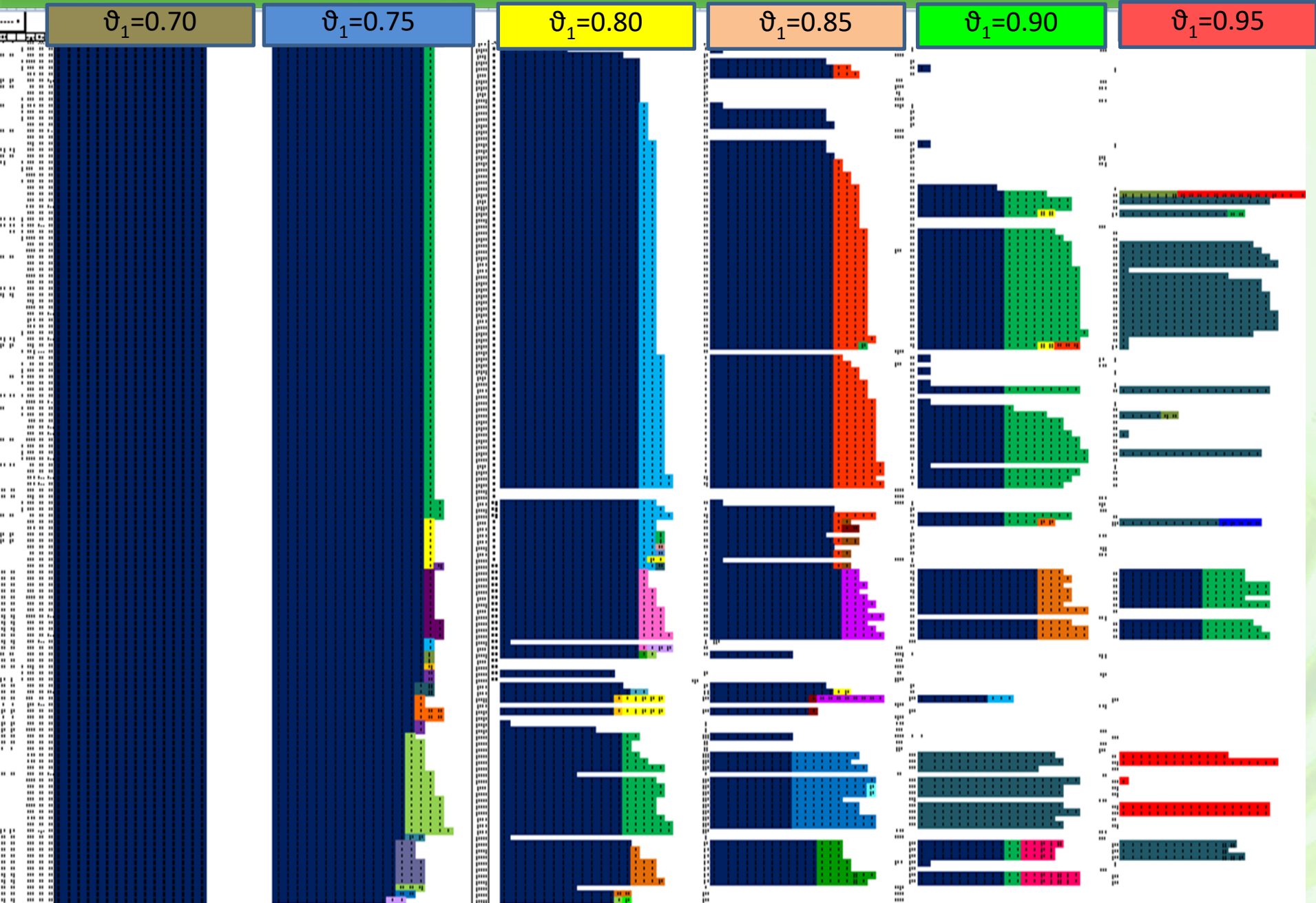
$\vartheta_2 = 0.0$

$\vartheta_{2/gene}$	0,0	0,1	0,5	1,0
a	1	1		
b	1	1	1	
c	1	1	1	1
d	1	1	1	1
e	1	1	1	
f	1			

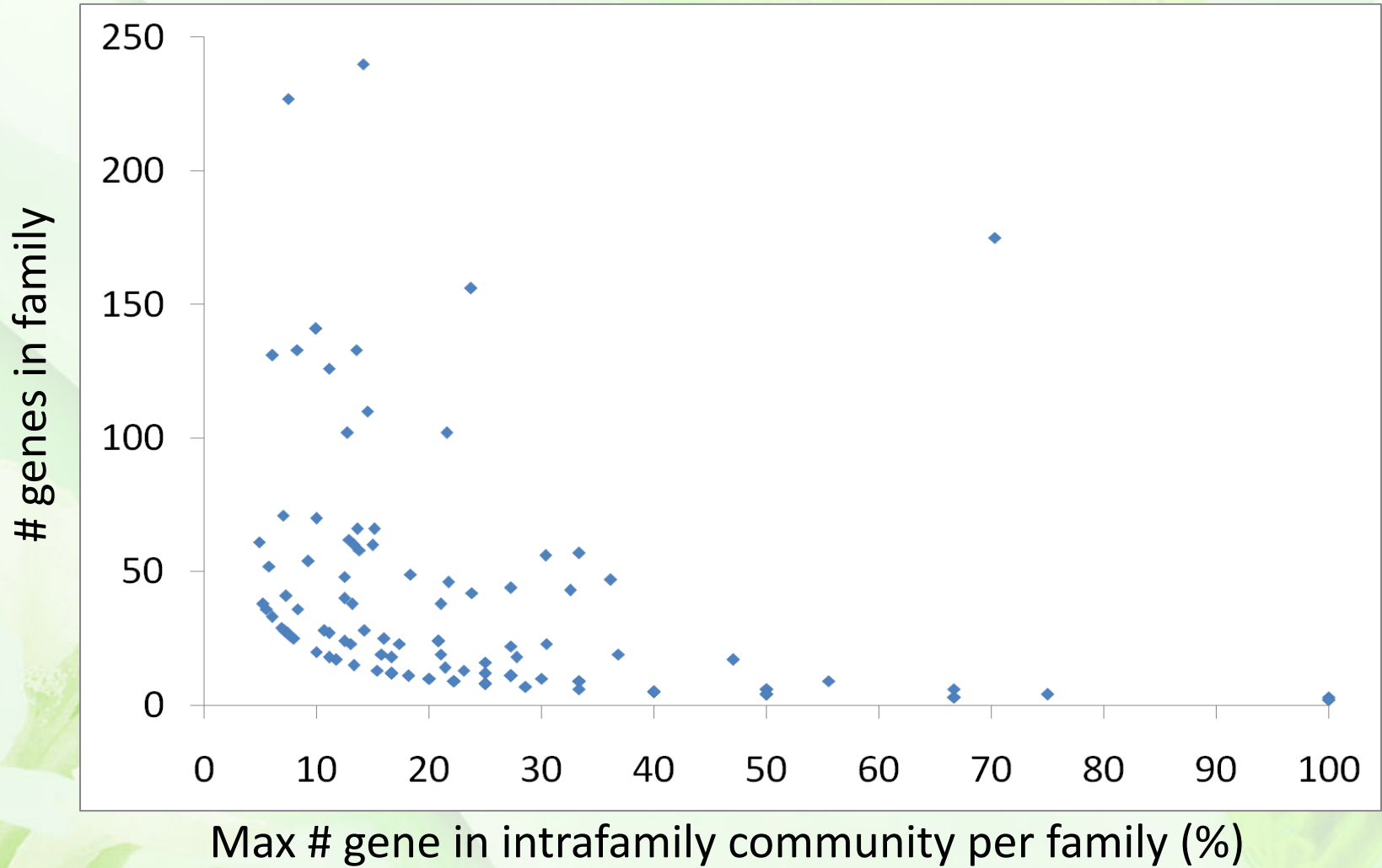
$\vartheta_1=0.70$ 

at_code	0.00	0.05	0.10	0.15	0.20	0.25	0.30	0.35	0.40	0.45	0.50	0.55	0.60	0.65	0.70	0.75	0.80	0.85	0.90	0.95	1.00	$\vartheta_2$	
AT3G16580	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1								
AT2G38010	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	8						
AT1G72280	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	8	12	9				
AT2G18420	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	8	12	9				
AT4G00040	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	8	12	9	14			
AT4G27540	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1								
AT3G18610																							
AT4G05240																							
AT4G31740																							
AT4G36150	1	1	1	1	1	1	1	1	1	1	1	1	1										
AT3G23010	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1								
AT3G13880	1	1	1	1	1	1	1	1	1	1	1	1											
AT5G25800	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	2						
AT3G62600	1	1	1	1	1	1	1	1	1	1	1	1	1	35	36	27	28	39	44	24	5		
AT4G16660	1	1	1	1	1	1	1	1	1	1	1	1	1	35	36	27	28	39	44	24	5		
AT5G45910	1	1	1	1	1	1	1	1	1	1	1	1	1	4									
AT4G07515	1	1	1	1	1	1	1	1	1	1	1	1	1	4	4	4							
AT2G17000	1	1	1	1	1	1	1	1	1	1	1	1	1	4	4	4	9						
AT3G04540	1	1	1	1	1	1	1	1	1	1	1	1	1	4	4	4	9	10					
AT1G69100	1	1	1	1	1	1	1	1	1	1	1	1	1	4	4	4	9	10					
AT1G47470	1	1	1	1	1	1	1	1	1	1	1	1	1	4	4	4	9						
AT5G42955	1	1	1	1	1	1	1	1	1	1	1	1	1	4	4	4	9	10					
AT5G64720	1	1	1	1	1	1	1	1	1	1	1	1	1	4	4	4	9	10	21				
AT2G21740																							
AT1G31450	1	1	1	1	1	1	1	1	1	1	1	1	1	4	4	4	9	10					
AT3G57840	1	1	1	1	1	1	1	1	1	1	1	1	1	4	4	4	9	10					
AT4G17505	1	1	1	1	1	1	1	1	1	1	1	1	1	4	4	4	9	10	21				
AT1G56620	1	1	1	1	1	1	1	1	1	1	1	1	1	4	4	4	9	10	21				
AT1G49530	1																						
AT2G16260	1	1	1	1	1	1	1																
AT1G19650	1	1	1	1	1	1	1	1	1	1	1	1	2										
AT2G40475	1	1	1	1	1	1	1	1	1	1	1	1	2	3									
AT1G74650	1																						
AT2G40200	1	1	1	1	1	1	1	1	1	1	1	1	1										
AT4G28160	1	1	1	1	1	1	1	1	1	1	1	1	2										
AT5G52390	1	1	1	1	1	1	1	1	1	1	1	1	2	3	10	10							
AT3G04520	1	1	1	1	1	1	1	1	1	1	1	1	1										
AT5G16080	1	1	1	1	1	1	1	1	1	1	1	1	1										
AT5G16120	1	1	1	1	1	1	1	1	1	1	1	1	1										
AT1G57650	1	1	1	1	1	1	1	1	1	1	1	1	1	5	9	8							
AT1G10585	1	1	1	1	1	1	1	1	1	1	1	1	1										
AT3G21890	1	1	1	1	1	1	1	1	1	1	1	1	1	5									
AT3G11480	1	1	1	1	1	1	1	1	1	1	1	1	1	5	9								
AT1G66570	1	1	1	1	1	1	1	1	1	1	1	1	1	5	9	8							
AT2G34810	1	1	1	1	1	1	1	1	1	1	1	1	1	5	9	8	14						

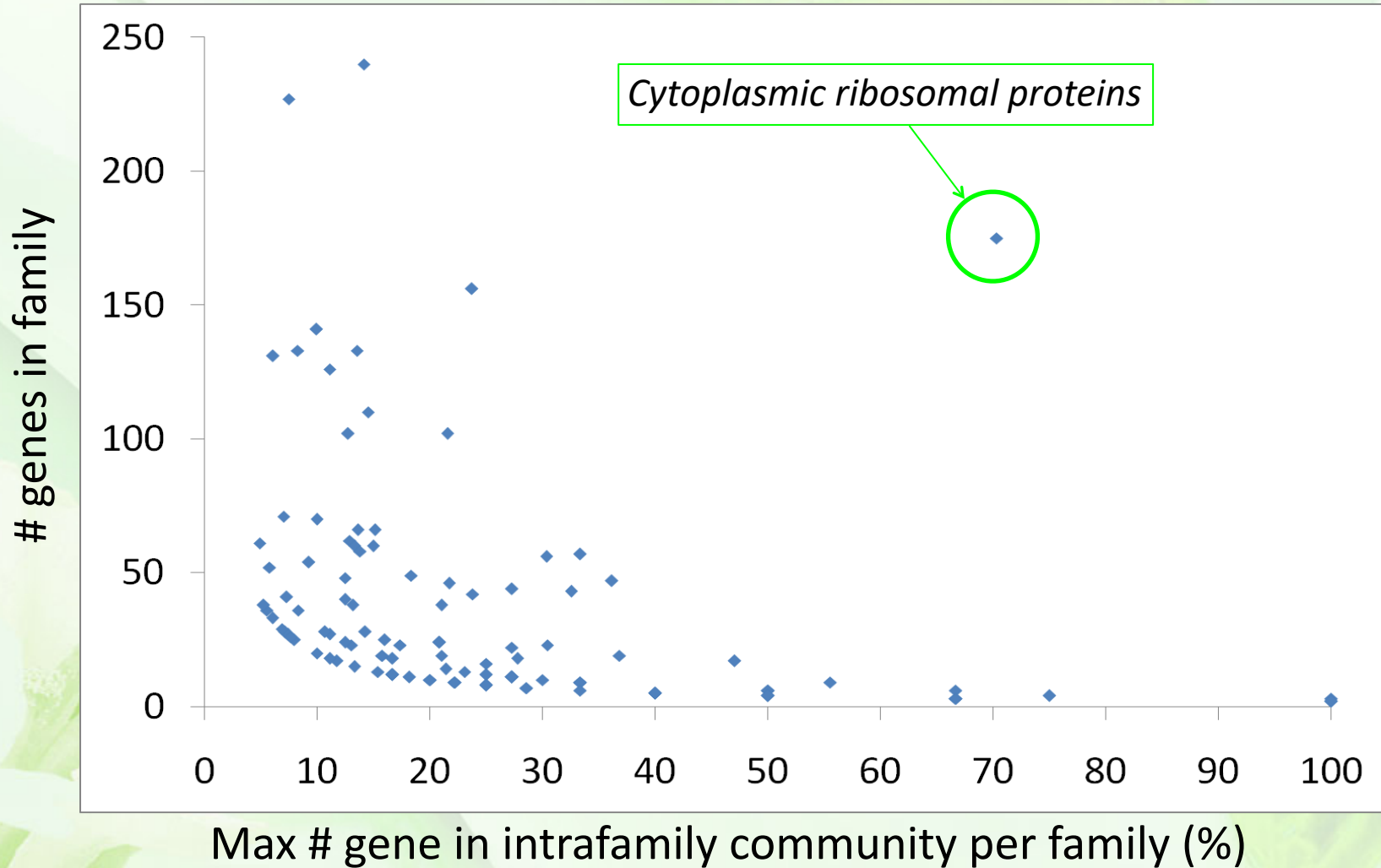
# Communities evolution through $\vartheta_2$ & $\vartheta_1$



# Gene families analysis



# Gene families analysis



# Θ1 = 0.95

θ2= 0.30	θ2= 0.35	θ2= 0.40	θ2= 0.45	θ2= 0.50	θ2= 0.55	θ2= 0.60	θ2= 0.65	θ2= 0.70	θ2= 0.75	θ2= 0.80	θ2= 0.85	θ2= 0.90	θ2= 0.95	Annotation	Function
5	5	5	5	5	6	6								60S acidic ribosomal pro	ribosomal components
5	5	5	5	5	6	6								60S acidic ribosomal pro	ribosomal components
5	5	5	5	5	6	6								Encodes cytoplasmic ribo	ribosomal components
5	5	5	5	5	6	6								60S ribosomal protein L3	ribosomal components
5	5	5	5	5	6	6								60S ribosomal protein L3	ribosomal components
5	5	5	5	5	6	6								60S ribosomal protein L3	ribosomal components
5	5	5	5	5	6	6								40S ribosomal protein S7	ribosomal components
5	5	5	5	5	6	6								ARABIDOPSIS THALIANA E	ribosomal components
5	5	5	5	5	6	6								60S ribosomal protein L1	ribosomal components
5	5	5	5	5	6	6								60S ribosomal protein L2	ribosomal components
5	5	5	5	5	6	6								RIBOSOMAL PROTEIN L23A	ribosomal components
5	5	5	5	5	6	6	7							acidic protein associate	ribosomal components
5	5	5	5	5	6	6	7							40S ribosomal protein S1	ribosomal components
5	5	5	5	5	6	6	7							60S acidic ribosomal pro	ribosomal components
5	5	5	5	5	6	6	7							40S ribosomal protein S3	ribosomal components
5	5	5	5	5	6	6	7							40S ribosomal protein S4	ribosomal components
5	5	5	5	5	6	6	7							60S ribosomal protein L7	ribosomal components
5	5	5	5	5	6	6	7	8						40S ribosomal protein S9	ribosomal components
5	5	5	5	5	6	6	7	8						ARABIDOPSIS RIBOSOMAL	ribosomal components
5	5	5	5	5	6	6	7	8						Encodes a protein with s	ribosomal components
5	5	5	5	5	6	6	7	8						60S ribosomal protein L1	ribosomal components
5	5	5	5	5	6	6	7	8						40S ribosomal protein S3	ribosomal components
5	5	5	5	5	6	6	7	8						40S ribosomal protein S2	ribosomal components
5	5	5	5	5	6	6	7	8						40S ribosomal protein S7	ribosomal components
5	5	5	5	5	6	6	7	8						40S ribosomal protein S1	ribosomal components
5	5	5	5	5	6	6	7	8	10					40S ribosomal protein S1	ribosomal components
5	5	5	5	5	6	6	7	8	10	11				unknown protein;	
5	5	5	5	5	6	6	7	8	10	11				RIBOSOMAL PROTEIN S11-	ribosomal components
5	5	5	5	5	6	6	7	8	10	11				40S ribosomal protein S1	ribosomal components
5	5	5	5	5	6	6	7	8	10	11				60S ribosomal protein L1	ribosomal components
5	5	5	5	5	6	6	7	8	23					60S ribosomal protein L4	ribosomal components
5	5	5	5	5	6	6	7	8	23					60S ribosomal protein L4	ribosomal components
5	5	5	5	5	6	6	7	8	23					60S ribosomal protein L5	ribosomal components
5	5	5	5	5	6	6	7	8	42					40S ribosomal protein S2	ribosomal components
5	5	5	5	5	6	6	7	8	42					60S ribosomal protein L2	ribosomal components
5	5	5	5	5	6	6	63							embryo defective 2386 (e	ribosomal components
5	5	5	5	5	6	6	63							RIBOSOMAL PROTEIN L23A	ribosomal components
5	5	5	5	5	6	6	68							40S ribosomal protein S1	ribosomal components
5	5	5	5	5	6	6	68							embryo defective 3010 (E	ribosomal components
5	5	5	5	5	6	41								60S ribosomal protein L3	ribosomal components
5	5	5	5	5	6	41								60S ribosomal protein L1	ribosomal components
5	5	5	40	45	40	49	59							60S ribosomal protein L3	ribosomal components
5	5	5	40	45	40	49	59							60S ribosomal protein L2	ribosomal components









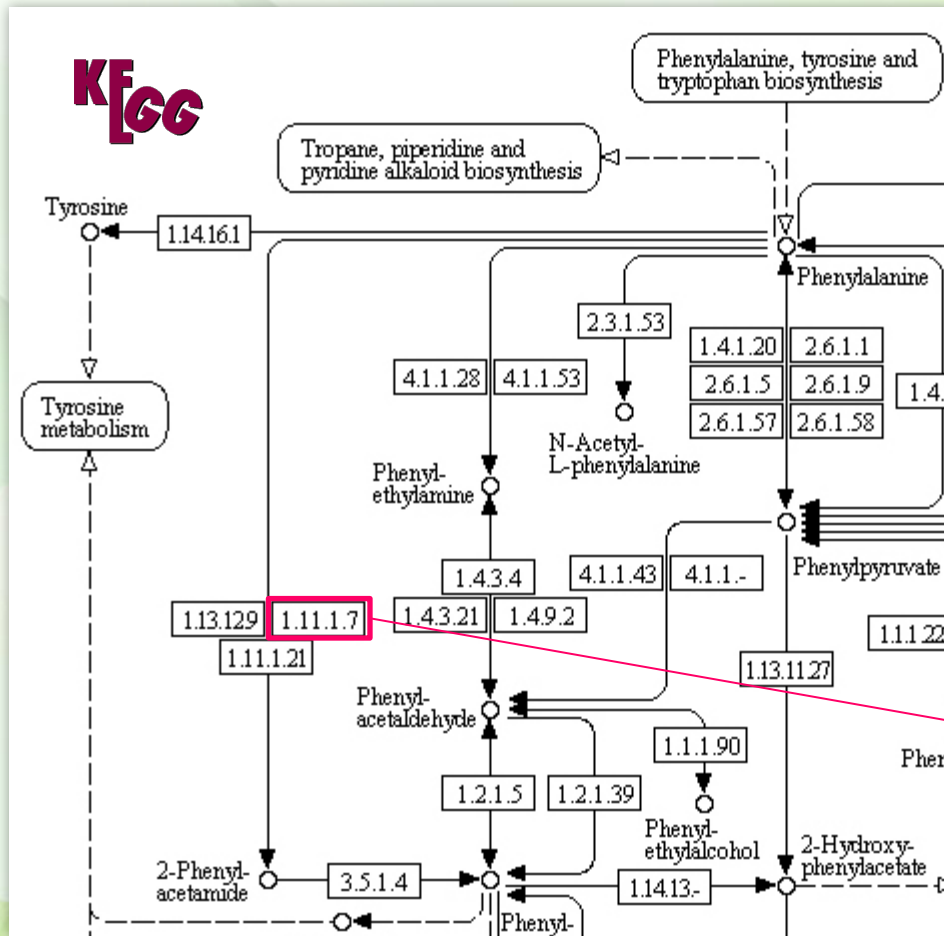
# Alcohol-forming Fatty Acyl-Coenzyme A Reductase

Genes in an exclusive network of paralogs:

		Genes			
Superfamily	Family	AT3G11980	AT3G44550	AT3G56700	AT3G44540
Transcription Factors	bHLH	Yellow	Blue	Grey	Grey
	bZIP	Grey	Blue	Green	Grey
	C2C2-CO-like	Grey	Blue	Green	Grey
	MYB	Yellow	Blue	Grey	Grey
	GRAS	Grey	Blue	Grey	Grey
	AP2-EREBP	Grey	Blue	Grey	Grey
	MADS	Grey	Blue	Grey	Grey
	NAC	Grey	Blue	Grey	Grey
WRKY	Grey	Blue	Grey	Grey	
ABC transporters	NAP subfamily	Grey	Blue	Grey	Grey
	WBC subfamily	Yellow	Grey	Grey	Grey
Acyl Lipidic Metabolism	Lipid Transfer Protein type 3	Yellow	Blue	Grey	Grey
	ELO-like Elongase	Grey	Blue	Grey	Grey
	Fatty Acid omega Hydrolase	Yellow	Grey	Grey	Grey
	Ketoacyl-CoA Synthase	Yellow	Grey	Grey	Grey
	Lipid Transfer Protein type 1	Yellow	Grey	Grey	Grey
	Lipid Transfer Protein type 5	Yellow	Grey	Grey	Grey
	Lipid Transfer Protein type 7	Grey	Grey	Grey	Pink
	Phosphatidyl Inisitol	Grey	Blue	Grey	Grey
	Plastidial Ketoacyl-ACP	Grey	Blue	Grey	Grey
	Transcription Factor CER2	Yellow	Grey	Grey	Grey
Other	Cytochrome P450	Yellow	Blue	Grey	Grey
	Glicoside Hydrolase	Grey	Blue	Grey	Grey
	Glicoside Transferase	Yellow	Blue	Grey	Grey
	Inorganic Solute Cotransporters	Yellow	Blue	Grey	Grey
	Miscellaneous Membrane Protein	Yellow	Blue	Grey	Grey
	Molignol Biosynthesis	Yellow	Blue	Grey	Grey
	Receptorkinase-like	Yellow	Blue	Grey	Grey
	Expansins	Grey	Blue	Grey	Grey
	Glutathione S-transferase	Yellow	Blue	Grey	Grey
Unknown	Yellow	Blue	Green	Pink	

# Pathway analysis

## Phenylalanine Metabolism from Kegg



Ref. Kegg <http://www.genome.jp/map00360>

Phenylalanine

2-

Phenylacetamide

1 Function...

43 Peroxidases described!

AT1G05240, AT1G05250, AT1G05260 (RCI3), AT1G14540, AT1G14550, AT1G24110, AT1G30870, AT1G34510, AT1G44970, AT1G48130 (PER1), AT1G49570, AT1G68850, AT1G71695, AT1G77100, AT2G18140, AT2G18150, AT2G18980, AT2G22420, AT2G24800, AT2G34060, AT2G35380, AT2G37130, AT2G38380, AT2G38390, AT2G39040, AT2G41480, AT2G43480, AT3G01190, AT3G03670, AT3G17070, AT3G21770, AT3G28200, AT3G32980, AT3G49110 (PRXCA), AT3G49120 (PRXCB), AT3G49960, AT3G50990, AT4G08770, AT4G08780, AT4G11290, AT4G16270, AT4G17690, AT4G21960 (PRXR1), A



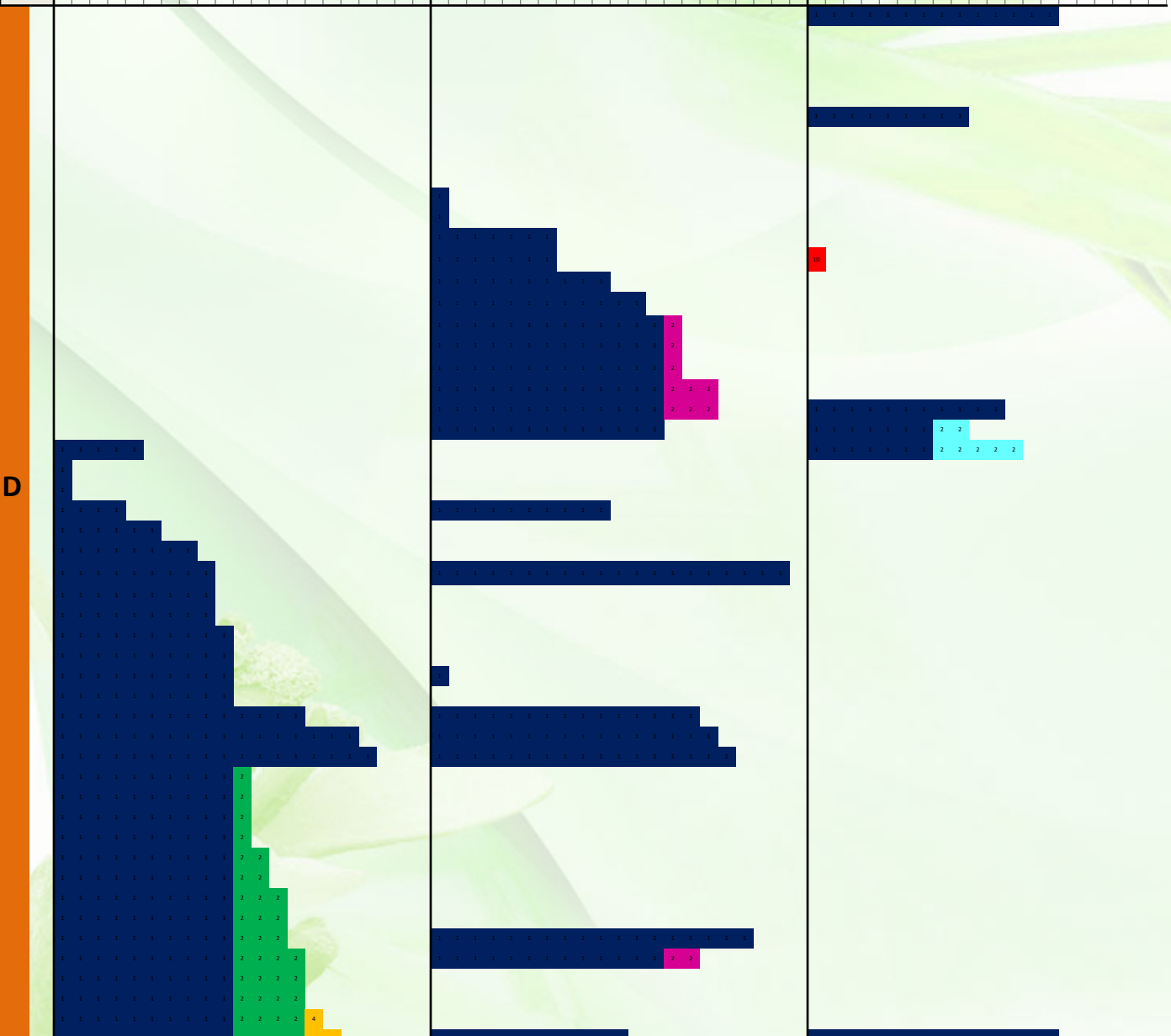
$\vartheta_1=0.90$

Whole Plant

Flower

Leaf

$\vartheta_2$  0.00 0.05 0.10 0.15 0.20 0.25 0.30 0.35 0.40 0.45 0.50 0.55 0.60 0.65 0.70 0.75 0.80 0.85 0.90 0.95 1.00 0.00 0.05 0.10 0.15 0.20 0.25 0.30 0.35 0.40 0.45 0.50 0.55 0.60 0.65 0.70 0.75 0.80 0.85 0.90 0.95 1.00 0.00 0.05 0.10 0.15 0.20 0.25 0.30 0.35 0.40 0.45 0.50 0.55 0.60 0.65 0.70 0.75 0.80 0.85 0.90 0.95 1.00



D

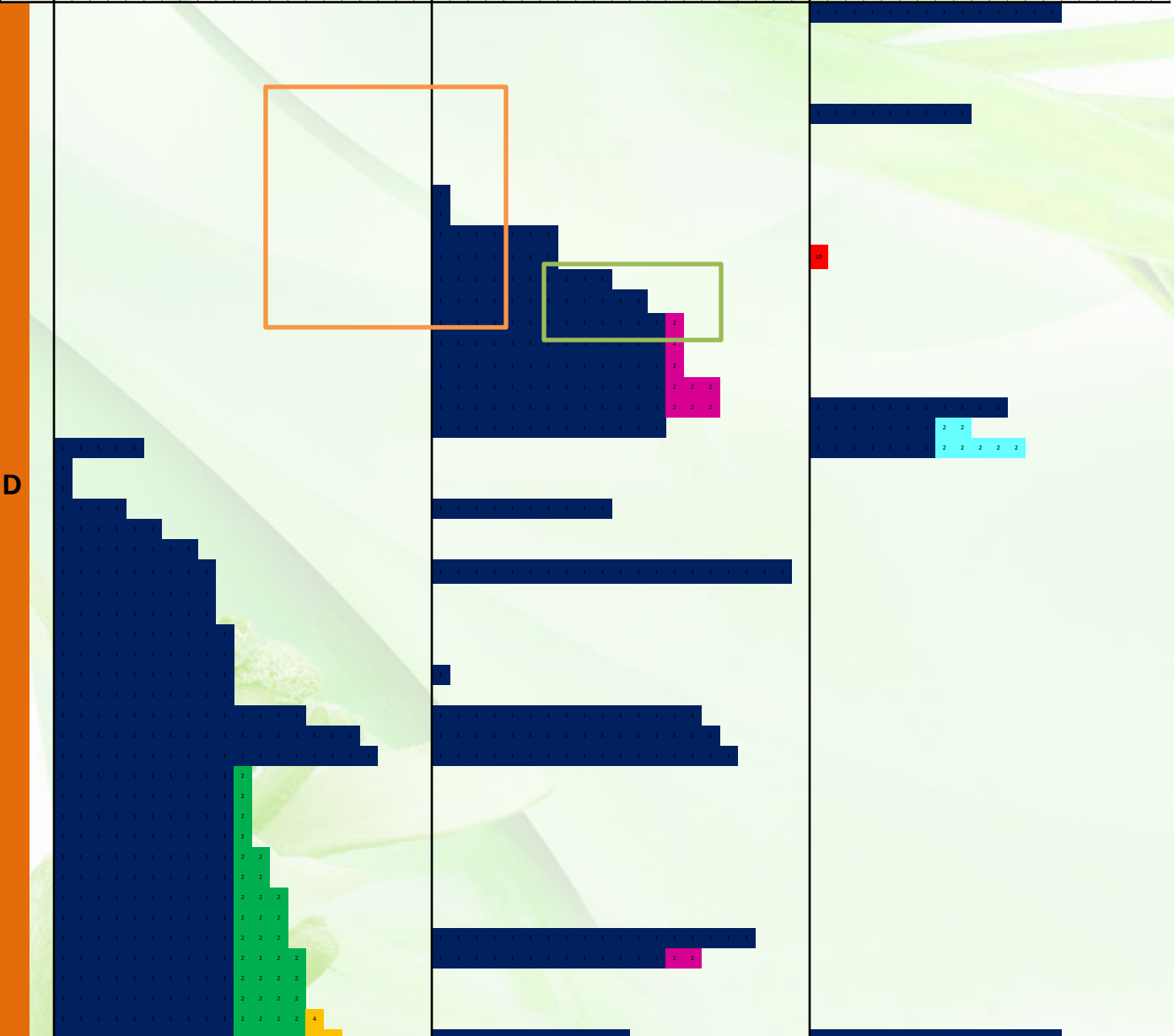
$\vartheta_1=0.90$

Whole Plant

Flower

Leaf

$\vartheta_2$  0.00 0.05 0.10 0.15 0.20 0.25 0.30 0.35 0.40 0.45 0.50 0.55 0.60 0.65 0.70 0.75 0.80 0.85 0.90 0.95 1.00 0.00 0.05 0.10 0.15 0.20 0.25 0.30 0.35 0.40 0.45 0.50 0.55 0.60 0.65 0.70 0.75 0.80 0.85 0.90 0.95 1.00 0.00 0.05 0.10 0.15 0.20 0.25 0.30 0.35 0.40 0.45 0.50 0.55 0.60 0.65 0.70 0.75 0.80 0.85 0.90 0.95 1.00





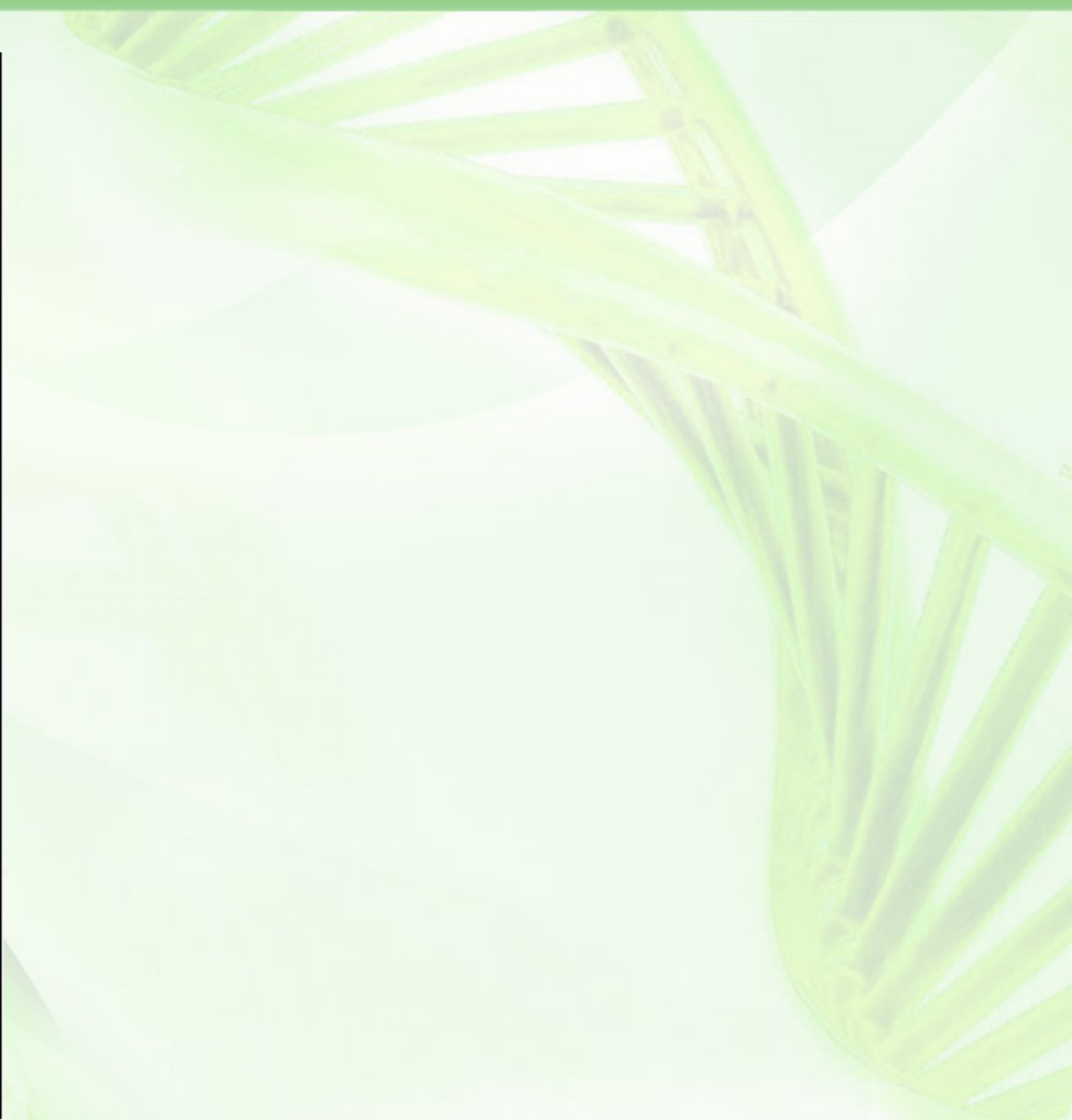
Microtubule  
Cytoskeleton  
Organization  
(GO:0000226)

146 genes:

-105 without  
families or in  
the same  
paralogs  
network

G.O.

Microtubule Cytoskeleton Organization (GO:0000226)	
AT1G03780	
AT1G08560	
AT1G15570	
AT1G18370	
AT1G20090	
AT1G20570	
AT1G27920	
AT1G49040	
AT1G51740	
AT1G59540	
AT1G63640	
AT1G72250	
AT1G80245	
AT1G80260	
AT2G18390	
AT2G21390	
AT2G22610	
AT2G25060	
AT2G30800	
AT2G36200	
AT2G38720	
AT3G10310	
AT3G19590	
AT3G26560	
AT3G43610	
AT3G44050	
AT3G53760	
AT4G00695	
AT4G05190	
AT4G14150	
AT4G14330	
AT4G21270	
AT4G22860	
AT4G26660	
AT4G26760	
AT4G27180	
AT4G29360	
AT4G31840	
AT4G35020	
AT4G37490	
AT5G06680	
AT5G13000	
AT5G13840	
AT5G15510	
AT5G17410	
AT5G18580	
AT5G23910	
AT5G25090	
AT5G51600	
AT5G54670	
AT5G55520	
AT5G62250	

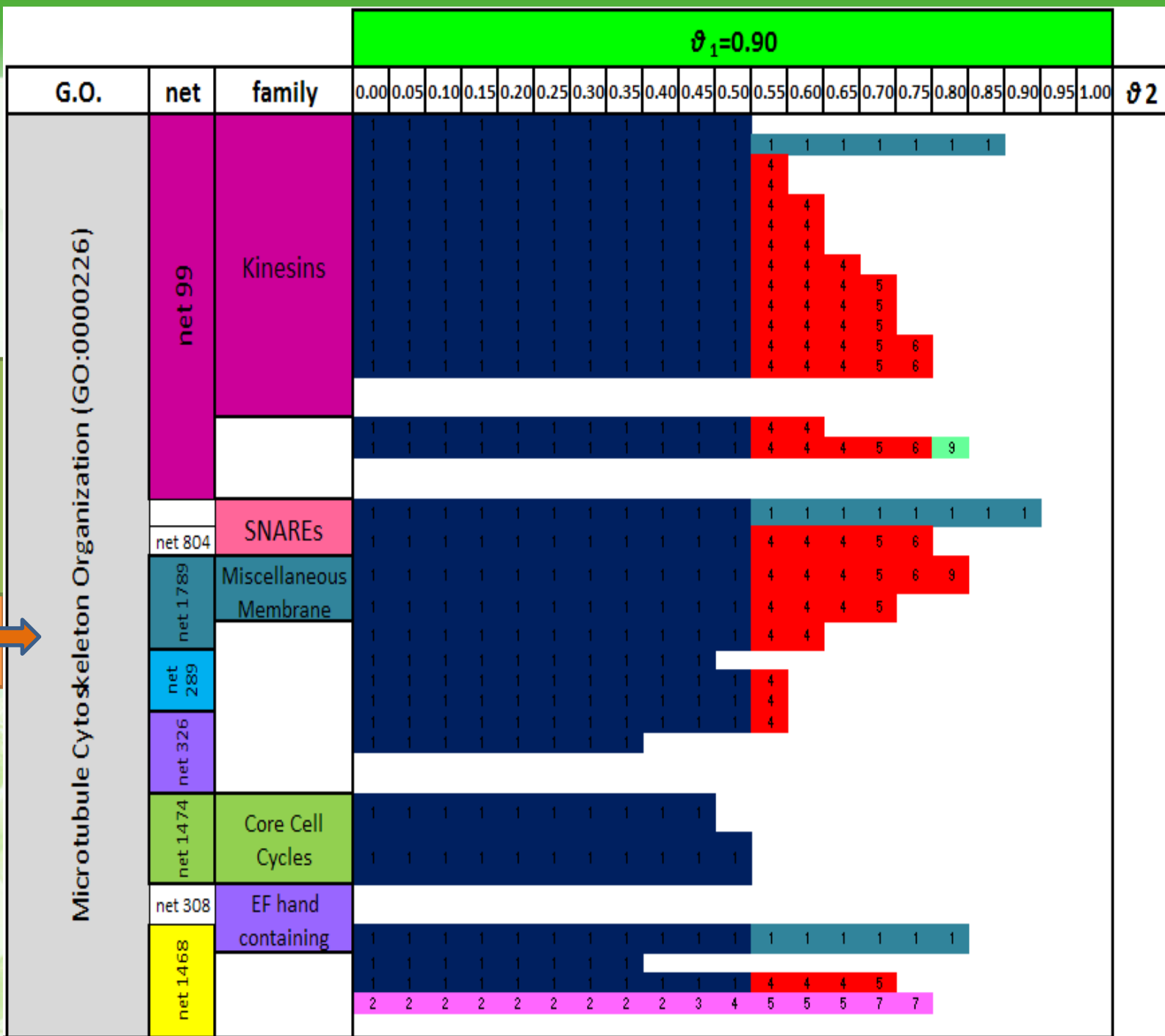


Microtubule Cytoskeleton Organization (GO:0000226)

146 genes:

-105 without families or in the same paralogs network

-41 with this distribution

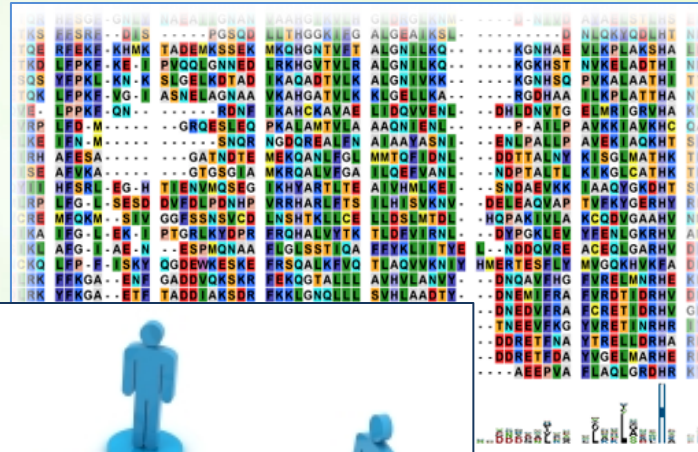
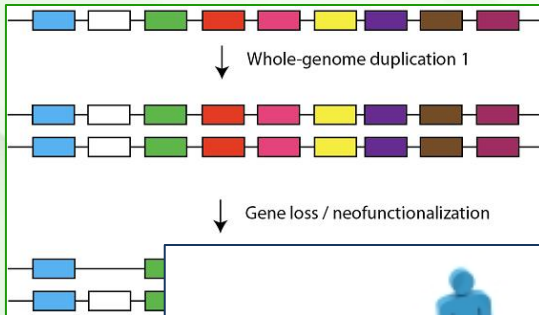


# Conclusions

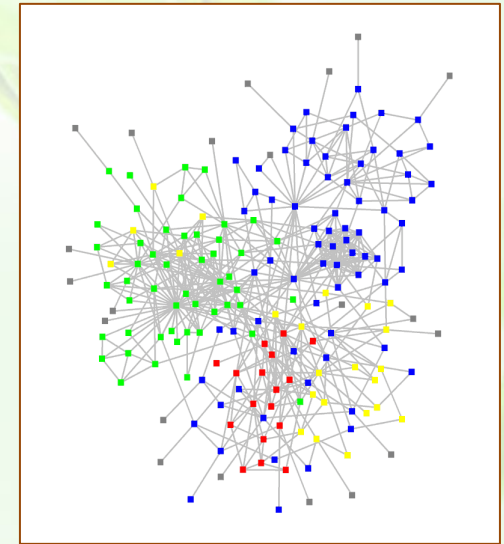
While a single dataset analysis may give a partial viewpoint of *A. thaliana* gene function, an exhaustive and integrated analysis on...

## Gene Families

### Paralogs



### Expression Data

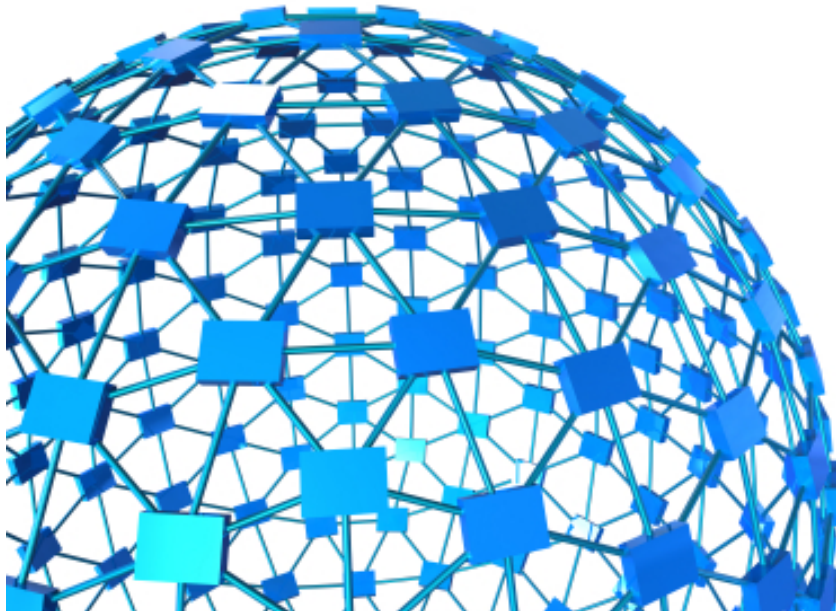


identifies the best target genes!

# Conclusions

## Future Aims:

- The consolidation of the clustering social network algorithm for gene characterization in *A. thaliana*
- A user friendly on line platform developed to data analysis integration and visualization of *A. thaliana* transcriptome



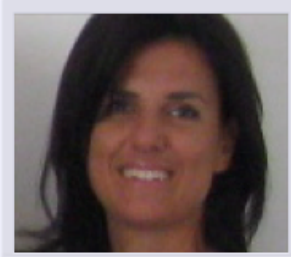
...work in progress



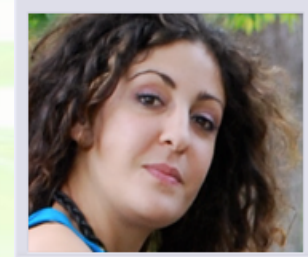
# Acknowledgements



**Prof. Luigi Frusciante**



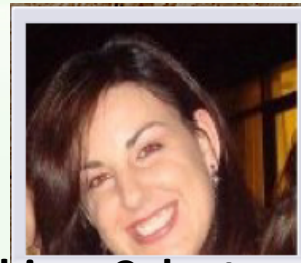
**Maria Luisa Chiusano**



**Alessandra Traini**



**Crescenzo Gallo**



**Chiara Colantuono**

# Thank You!

Contact: [pasquale.disalle@libero.it](mailto:pasquale.disalle@libero.it)  
PI: [chiusano@unina.it](mailto:chiusano@unina.it)