

Plant Taxonomic Network Projection

Chloe T. Calderon

Department of Physics and Astronomy - Northwestern University

Abstract

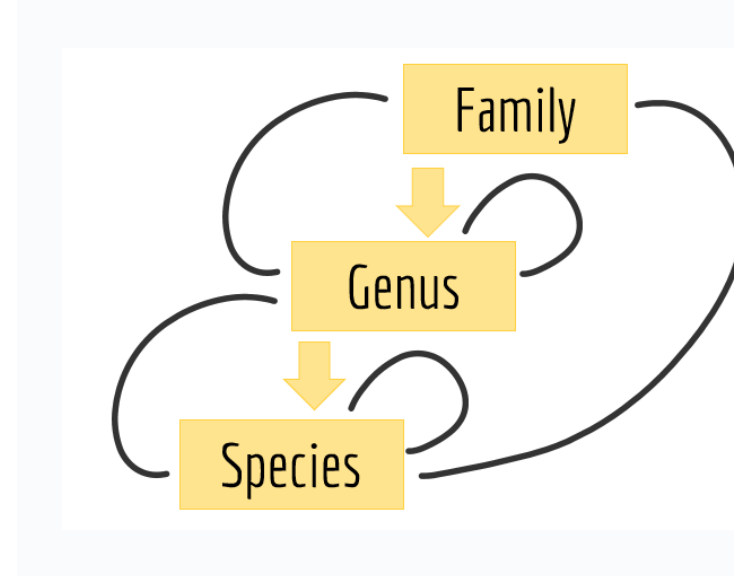
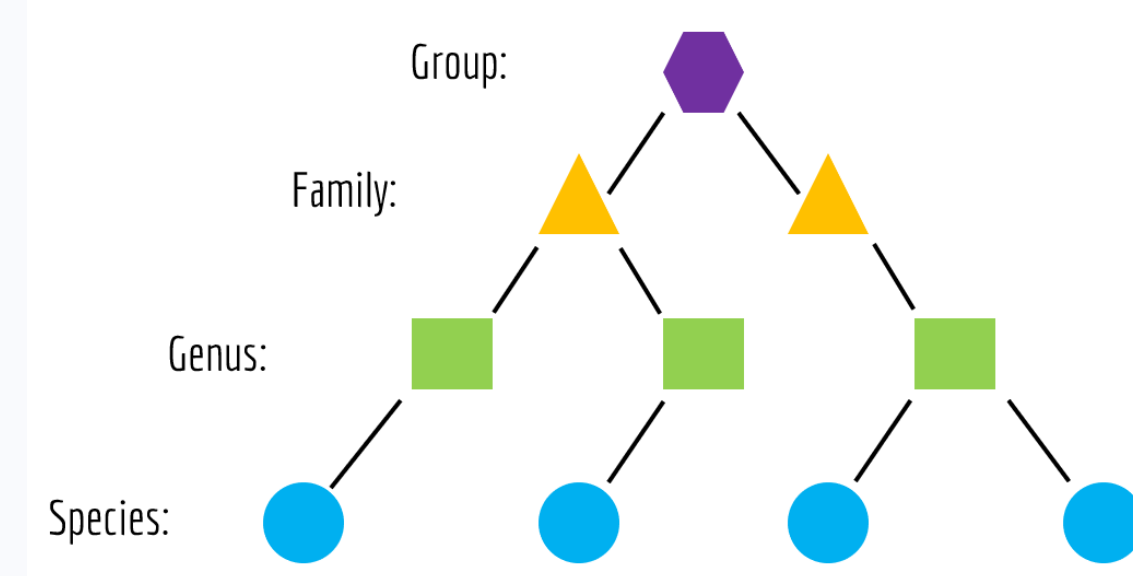
Plant taxonomic classification is based on observable, macroscopic shared characteristics.

What would a classification system based on genetic characteristics (number of genes, number of copies, etc.) look like? How do genetic characteristics relate to our current classification system?

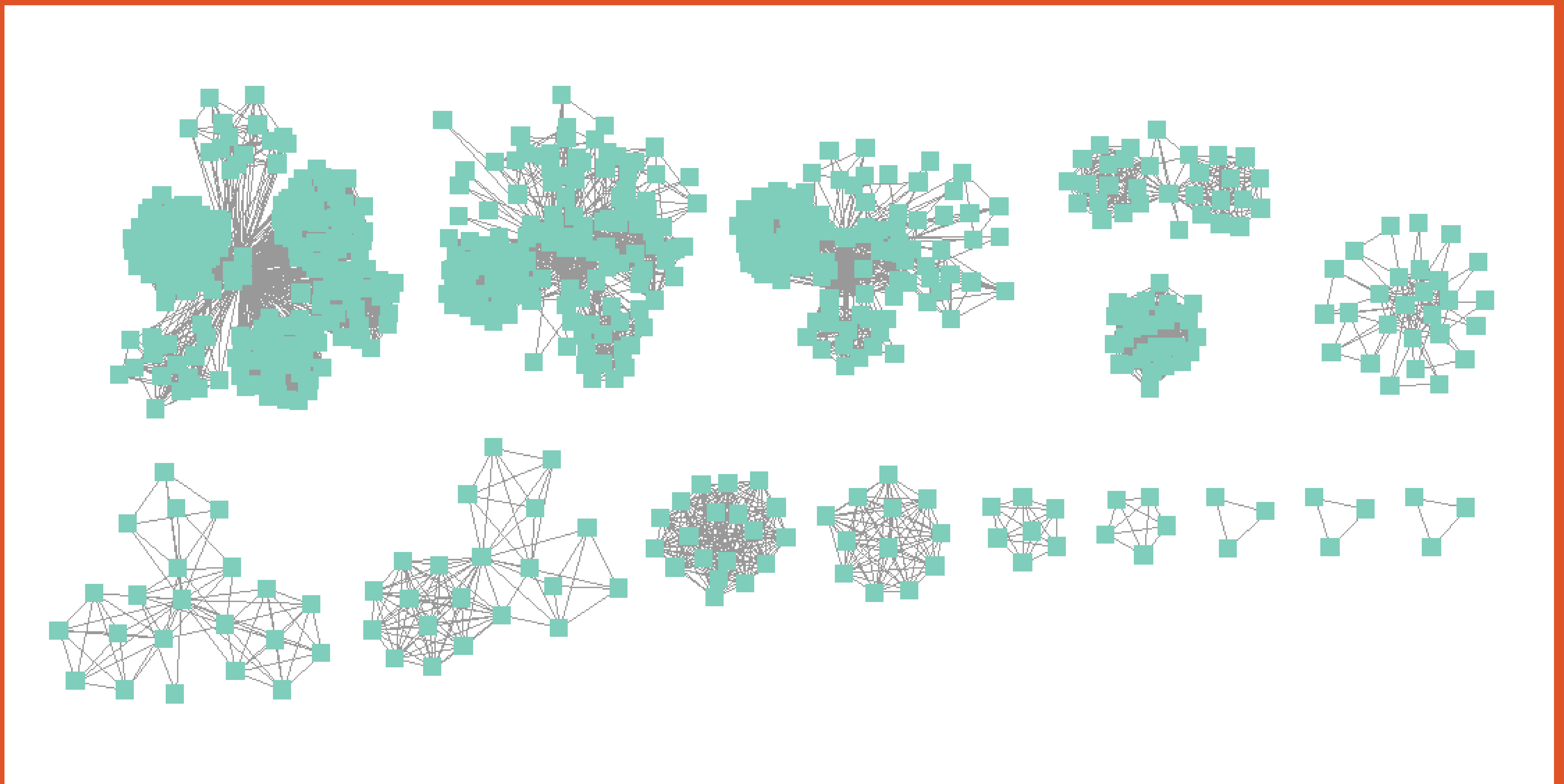
We begin to answer these questions studying the tripartite network for the Gymnosperm taxonomic group as a single network. The original network has a hierarchical structure that is difficult to gain information from, thus we will project this network. The three kinds of nodes for this tripartite network are the taxonomic family, genus, and species.

Method

We began with data from the *TRY Plant Trait Database*. There were 66,043 species from ~10 groups in the database. Our analysis was narrowed to one group, the Gymnosperms, plants with unprotected seeds including conifers, cycads, and ginkos. In this group, there are 15 families, 77 genera, and 647 species. Links were formed between species in the same genus, genera in the same family, species and its genera, genera and its family, and species and its family. Cytoscape was used to analyze this projection network.

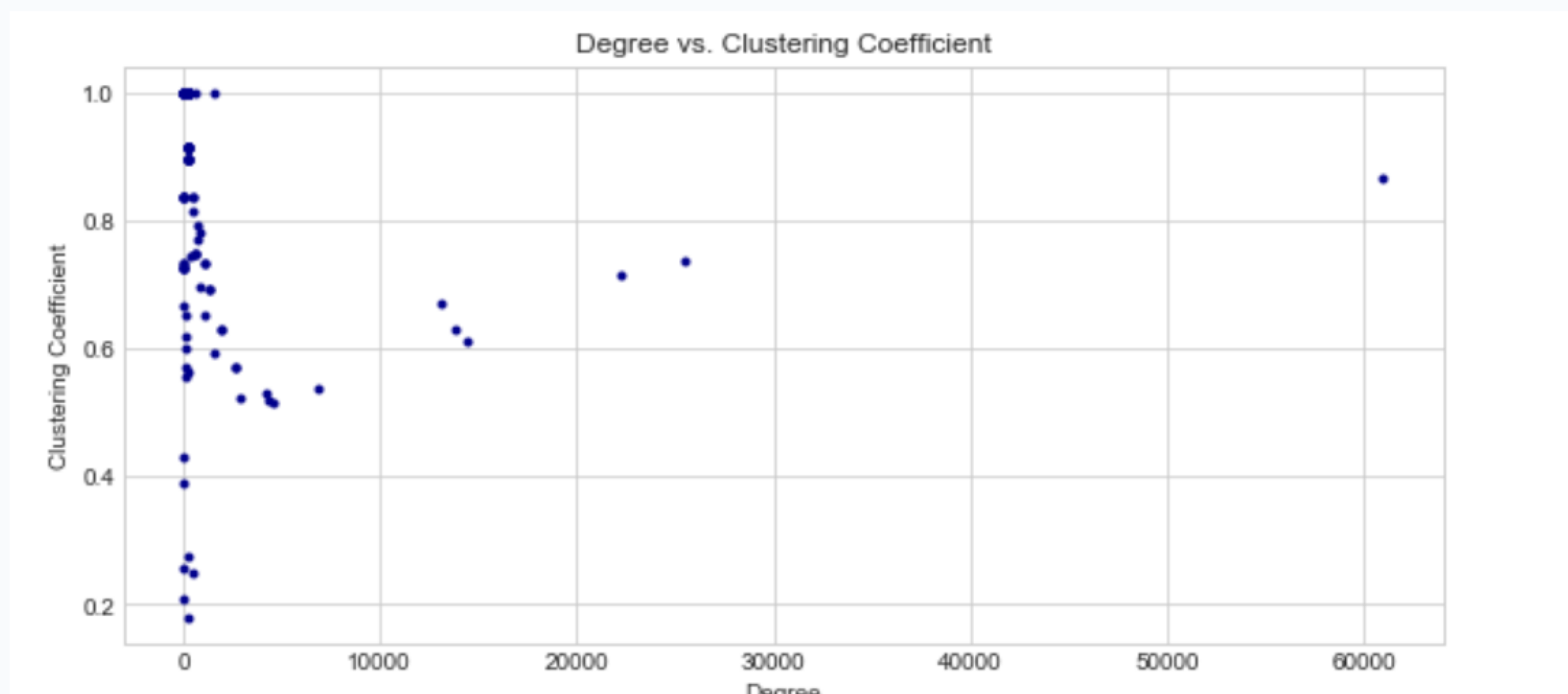
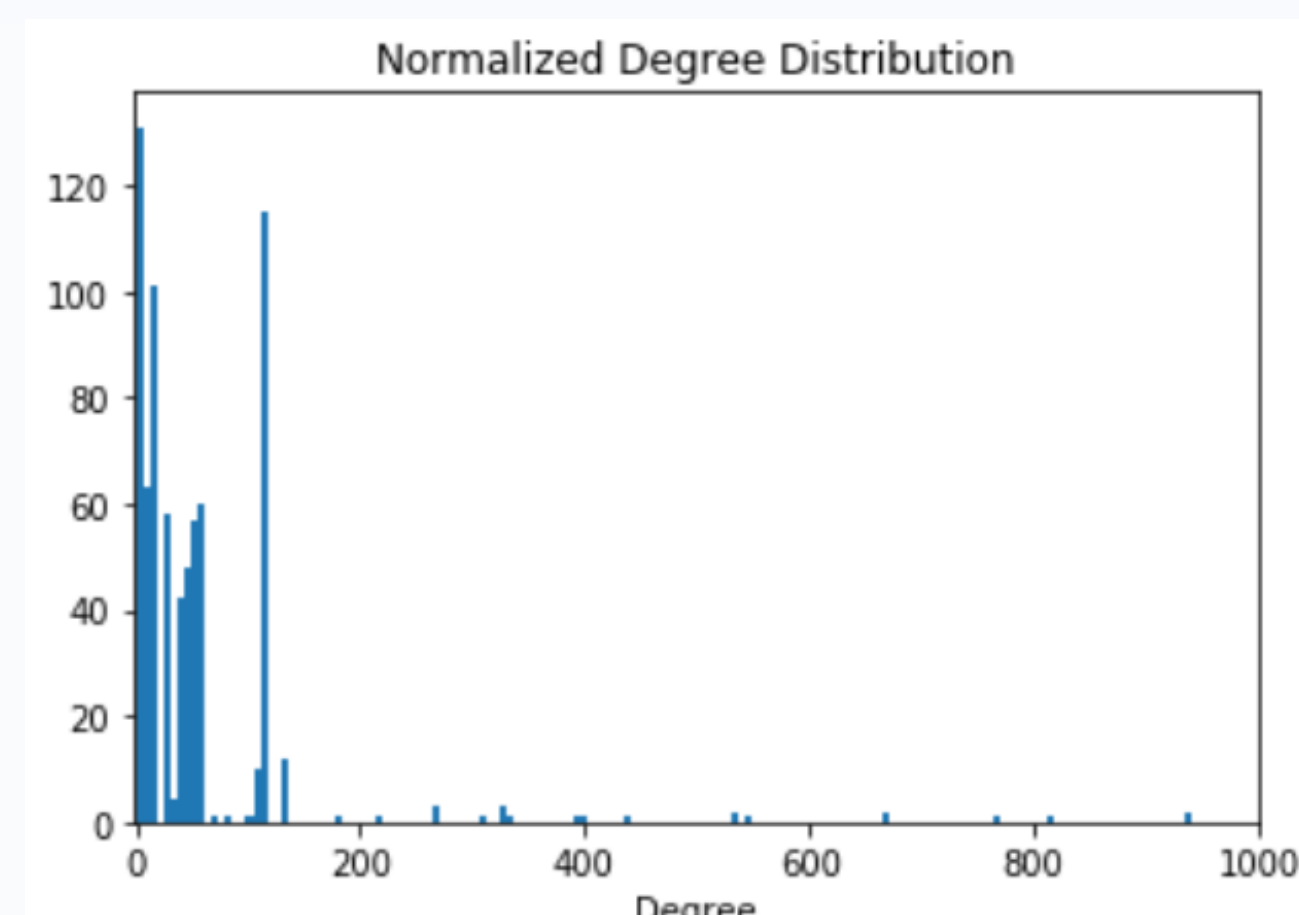


	D	E	F	G	H
1	Genus	Species/Ethet	Family	PhylogeneticGroup	PlantGrowthForm
2	Abrotia	maritima	Nyctaginaceae	Angiosperm_Eudicotyl	herb
3	Abutilon	theophrasti	Malvaceae	Angiosperm_Eudicotyl	herb
4	Acacia	anura	Fabaceae	Angiosperm_Eudicotyl	shrub/tree
5	Acacia	bivenata	Fabaceae	Angiosperm_Eudicotyl	shrub/tree
6	Acacia	busilini	Fabaceae	Angiosperm_Eudicotyl	shrub/tree
7	Acacia	cognata	Fabaceae	Angiosperm_Eudicotyl	shrub/tree
8	Acacia	deubata	Fabaceae	Angiosperm_Eudicotyl	shrub/tree
9	Acacia	elata	Fabaceae	Angiosperm_Eudicotyl	tree
10	Acacia	implexa	Fabaceae	Angiosperm_Eudicotyl	shrub/tree
11	Acacia	irrorata	Fabaceae	Angiosperm_Eudicotyl	shrub/tree
12	Acacia	jonesii	Fabaceae	Angiosperm_Eudicotyl	shrub
13	Acacia	koa	Fabaceae	Angiosperm_Eudicotyl	tree
14	Acacia	ligulata	Fabaceae	Angiosperm_Eudicotyl	shrub/tree
15	Acacia	mearnsii	Fabaceae	Angiosperm_Eudicotyl	tree
16	Acacia	melanoxylon	Fabaceae	Angiosperm_Eudicotyl	tree
17	Acacia	penulata	Fabaceae	Angiosperm_Eudicotyl	tree



Results

Number of Nodes: 739
 Number of Links: 131,321
 Connected Components: 15
 Clustering Coefficient: 0.986
 Smallest Degree: 2
 Lowest Degree: 60,950



Discussion

For our plant taxonomic network projection, the number of nodes is the sum of the families, genera, and species in the taxonomic group, Gymnosperms, thus is exactly what was expected. The number of possible links for 739 nodes follows $\frac{N(N-1)}{2} = 272,691$, whereas the number of links is 131,321. This value makes is what one might expect, since the nodes from different families (components) are by definition not able to form a link. There are 15 connected components, just what one might predict since there are 15 families in our dataset. The clustering coefficient is very close to 1. This is what one would expect for each component is very near fully connected. Since the network is unweighted, the degree distribution does not give us the most useful information.

Moving Forward

We should transform this unweighted, projected network to a weighted network. To as the original question, does the taxonomic classification correlate with genetic classifications, we need to compare the results of the analysis of the weighted network to a dataset with ground truth data on genetic characteristics.

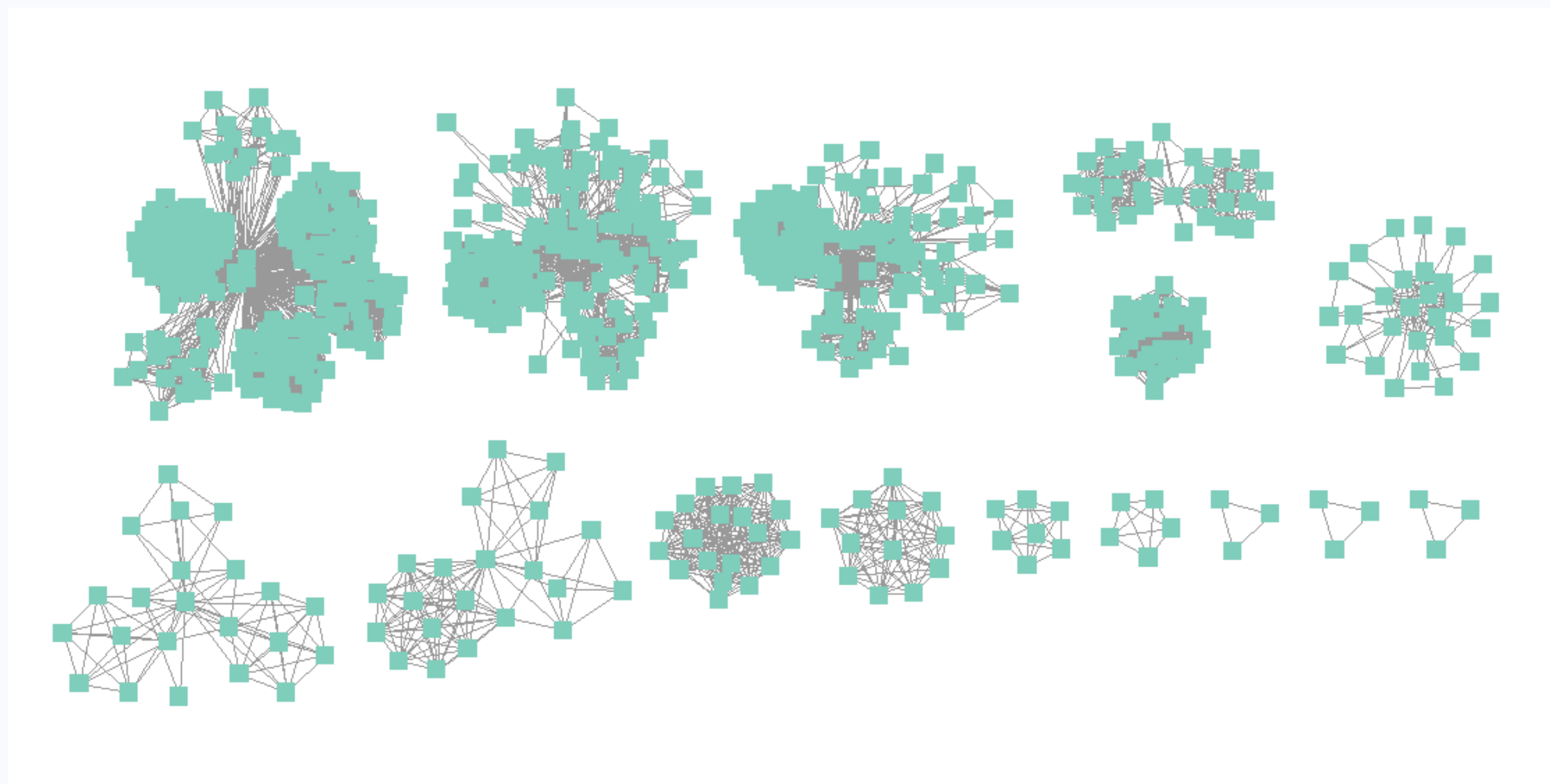
Acknowledgments

The *TRY* initiative and data base is hosted, developed and maintained by J. Kattge and G. Bönisch (Max Planck Institute for Biogeochemistry, Jena, Germany).
 Cytoscape Program

Plant Taxonomic Network Projection

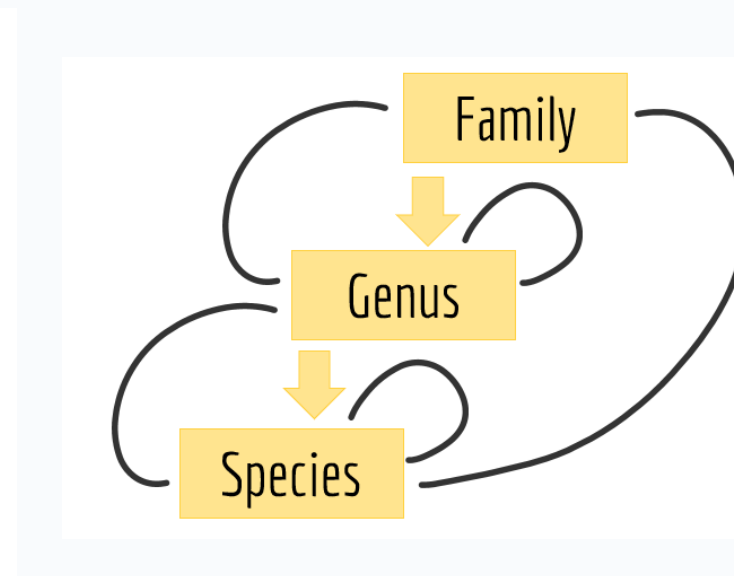
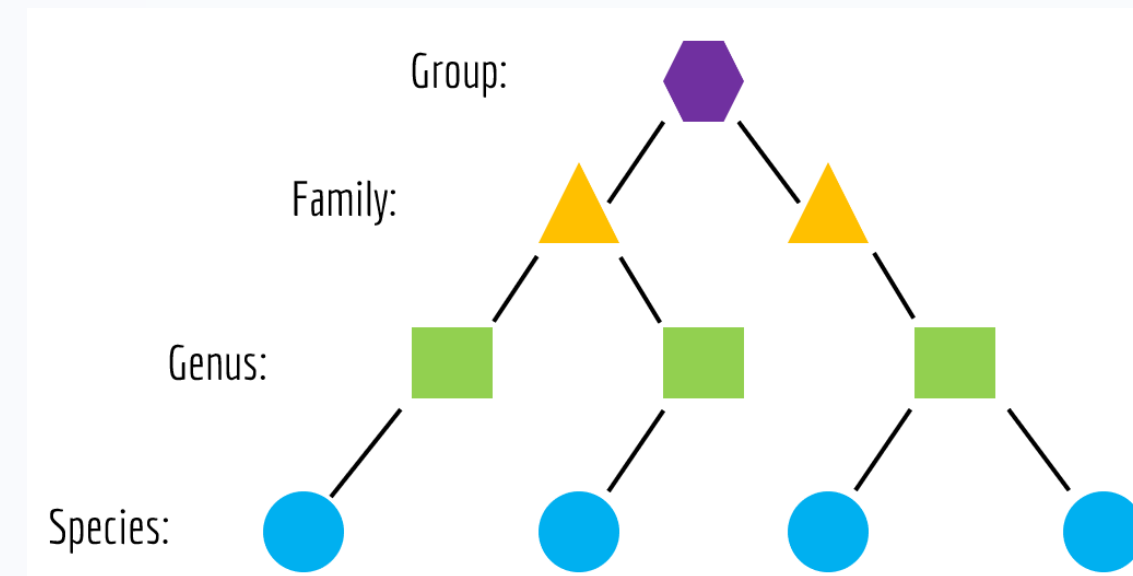
Chloe T. Calderon

Department of Physics and Astronomy - Northwestern University



Method

We began with data from the *TRY Plant Trait Database*. There were 66,043 species from ~10 groups in the database. Our analysis was narrowed to one group, the Gymnosperms, plants with unprotected seeds including conifers, cycads, and ginkos. In this group, there are 15 families, 77 genera, and 647 species. Links were formed between species in the same genus, genera in the same family, species and its genera, genera and its family, and species and its family. Cytoscape was used to analyze this projection network.



	D	E	F	G	H
1	Genus	Species/Ethet	Family	PhylogeneticGroup	PlantGrowthForm
2	Abrotia	maritima	Nyctaginaceae	Angiosperm_Eudicotyl	herb
3	Adiantum	theophrasti	Mahoeaceae	Angiosperm_Eudicotyl	herb
4	Acacia	anura	Fabaceae	Angiosperm_Eudicotyl	shrub/tree
5	Acacia	bivenata	Fabaceae	Angiosperm_Eudicotyl	shrub/tree
6	Acacia	buschii	Fabaceae	Angiosperm_Eudicotyl	shrub/tree
7	Acacia	cognata	Fabaceae	Angiosperm_Eudicotyl	shrub/tree
8	Acacia	deubata	Fabaceae	Angiosperm_Eudicotyl	shrub/tree
9	Acacia	elata	Fabaceae	Angiosperm_Eudicotyl	tree
10	Acacia	implexa	Fabaceae	Angiosperm_Eudicotyl	shrub/tree
11	Acacia	irrorata	Fabaceae	Angiosperm_Eudicotyl	shrub/tree
12	Acacia	jonesii	Fabaceae	Angiosperm_Eudicotyl	shrub
13	Acacia	koa	Fabaceae	Angiosperm_Eudicotyl	tree
14	Acacia	ligulata	Fabaceae	Angiosperm_Eudicotyl	shrub/tree
15	Acacia	mearnsii	Fabaceae	Angiosperm_Eudicotyl	tree
16	Acacia	melanoxylon	Fabaceae	Angiosperm_Eudicotyl	tree
17	Acacia	penulata	Fabaceae	Angiosperm_Eudicotyl	tree

Abstract

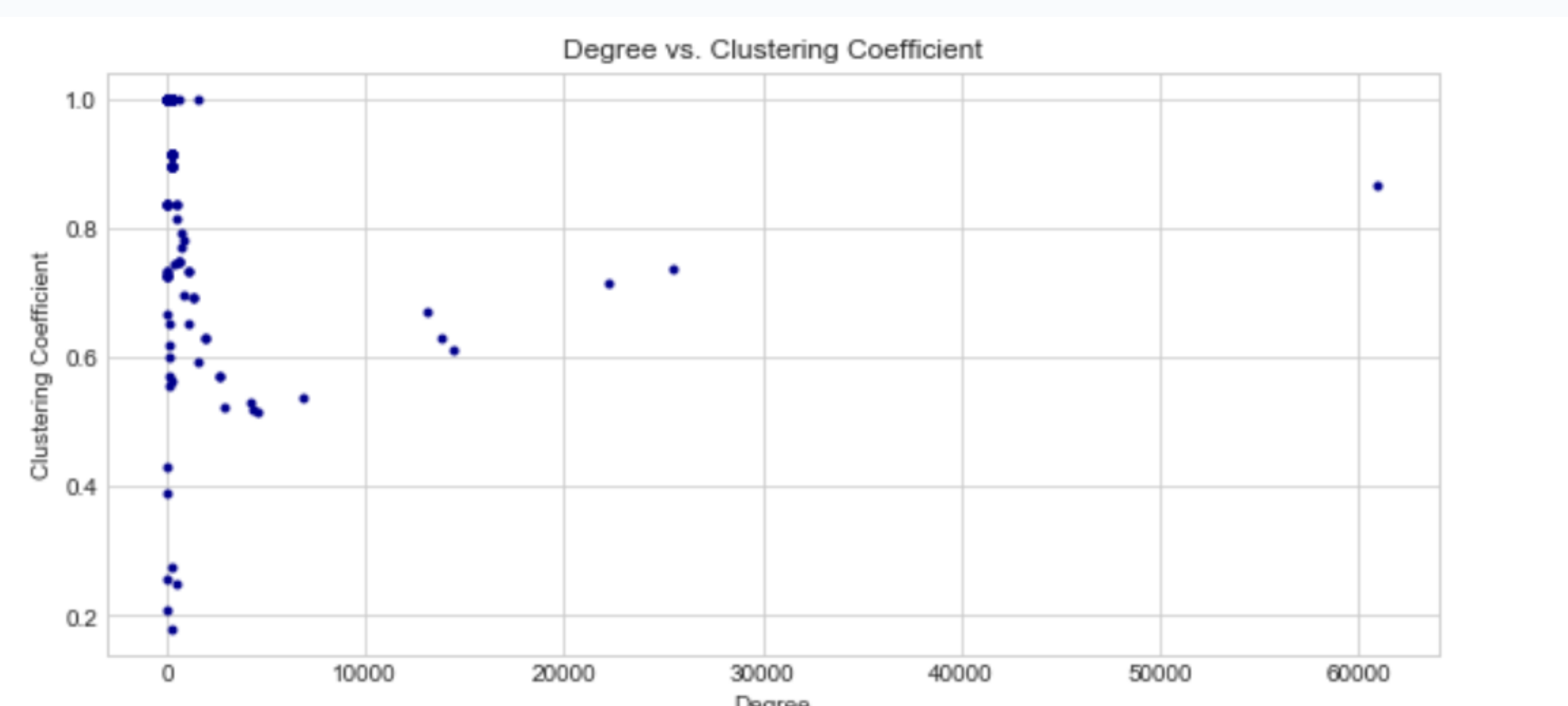
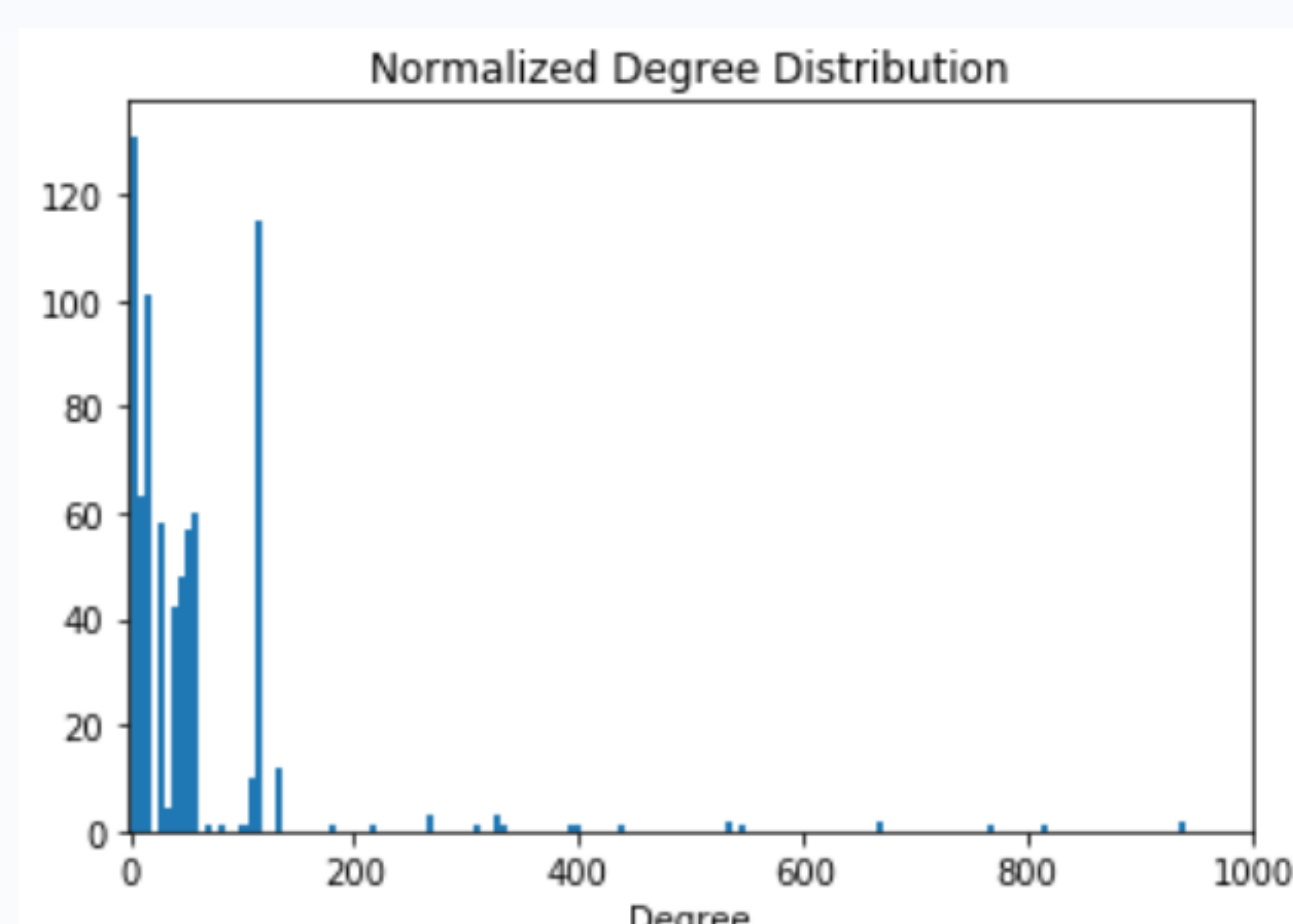
Plant taxonomic classification is based on observable, macroscopic shared characteristics.

What would a classification system based on genetic characteristics (number of genes, number of copies, etc.) look like? How do genetic characteristics relate to our current classification system?

We begin to answer these questions studying the tripartite network for the Gymnosperm taxonomic group as a single network. The original network has a hierarchical structure that is difficult to gain information from, thus we will project this network. The three kinds of nodes for this tripartite network are the taxonomic family, genus, and species.

Results

Number of Nodes: 739
Number of Links: 131,321
Connected Components: 15
Clustering Coefficient: 0.986
Smallest Degree: 2
Lowest Degree: 60,950



Discussion

For our plant taxonomic network projection, the number of nodes is the sum of the families, genera, and species in the taxonomic group, Gymnosperms, thus is exactly what was expected. The number of possible links for 739 nodes follows $\frac{N(N-1)}{2} = 272,691$, whereas the number of links is 131,321. This value makes is what one might expect, since the nodes from different families (components) are by definition not able to form a link. There are 15 connected components, just what one might predict since there are 15 families in our dataset. The clustering coefficient is very close to 1. This is what one would expect for each component is very near fully connected. Since the network is unweighted, the degree distribution does not give us the most useful information.

Moving Forward

We should transform this unweighted, projected network to a weighted network. To as the original question, does the taxonomic classification correlate with genetic classifications, we need to compare the results of the analysis of the weighted network to a dataset with ground truth data on genetic characteristics.

Acknowledgments

The *TRY* initiative and data base is hosted, developed and maintained by J. Kattge and G. Bönisch (Max Planck Institute for Biogeochemistry, Jena, Germany).
Cytoscape Program

Plant Taxonomic Network Projection

Chloe T. Calderon

Department of Physics and Astronomy - Northwestern University

Abstract

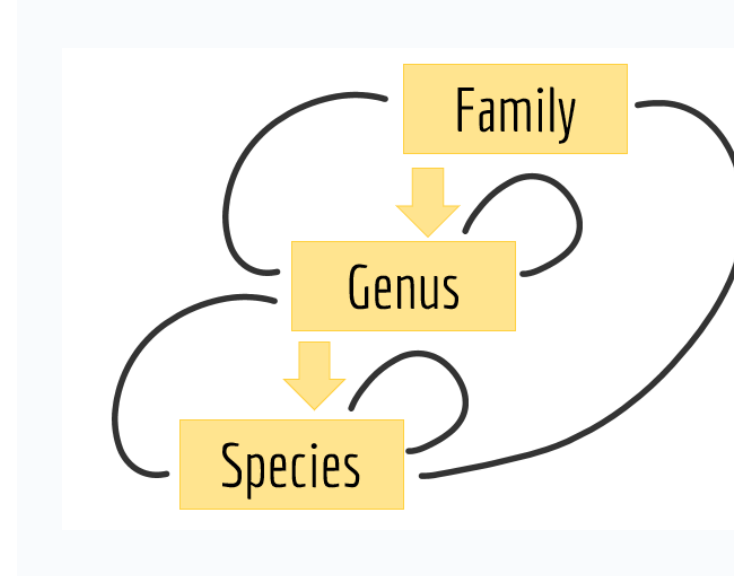
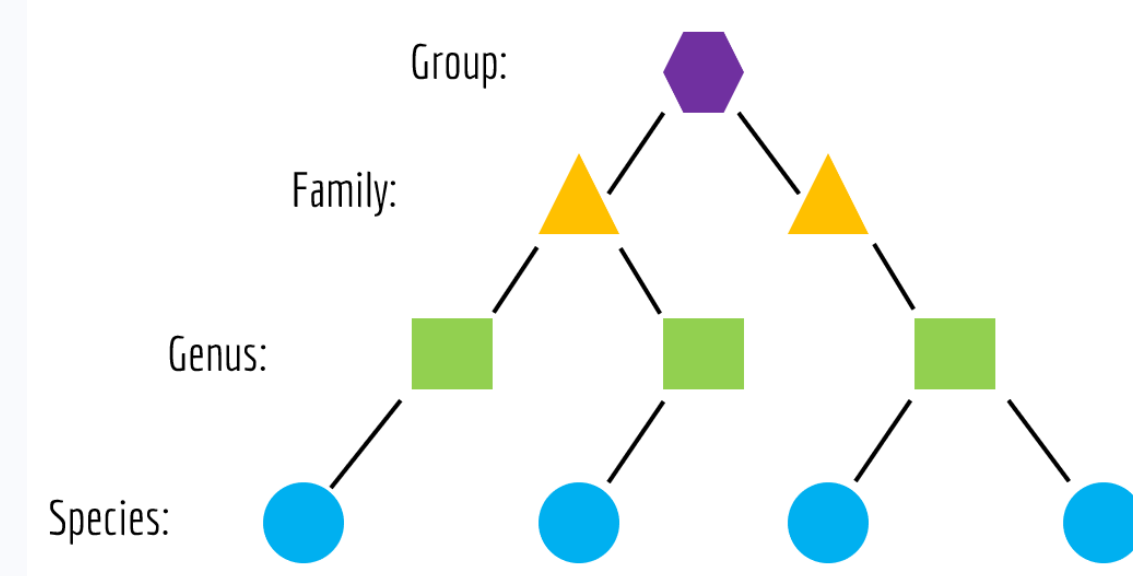
Plant taxonomic classification is based on observable, macroscopic shared characteristics.

What would a classification system based on genetic characteristics (number of genes, number of copies, etc.) look like? How do genetic characteristics relate to our current classification system?

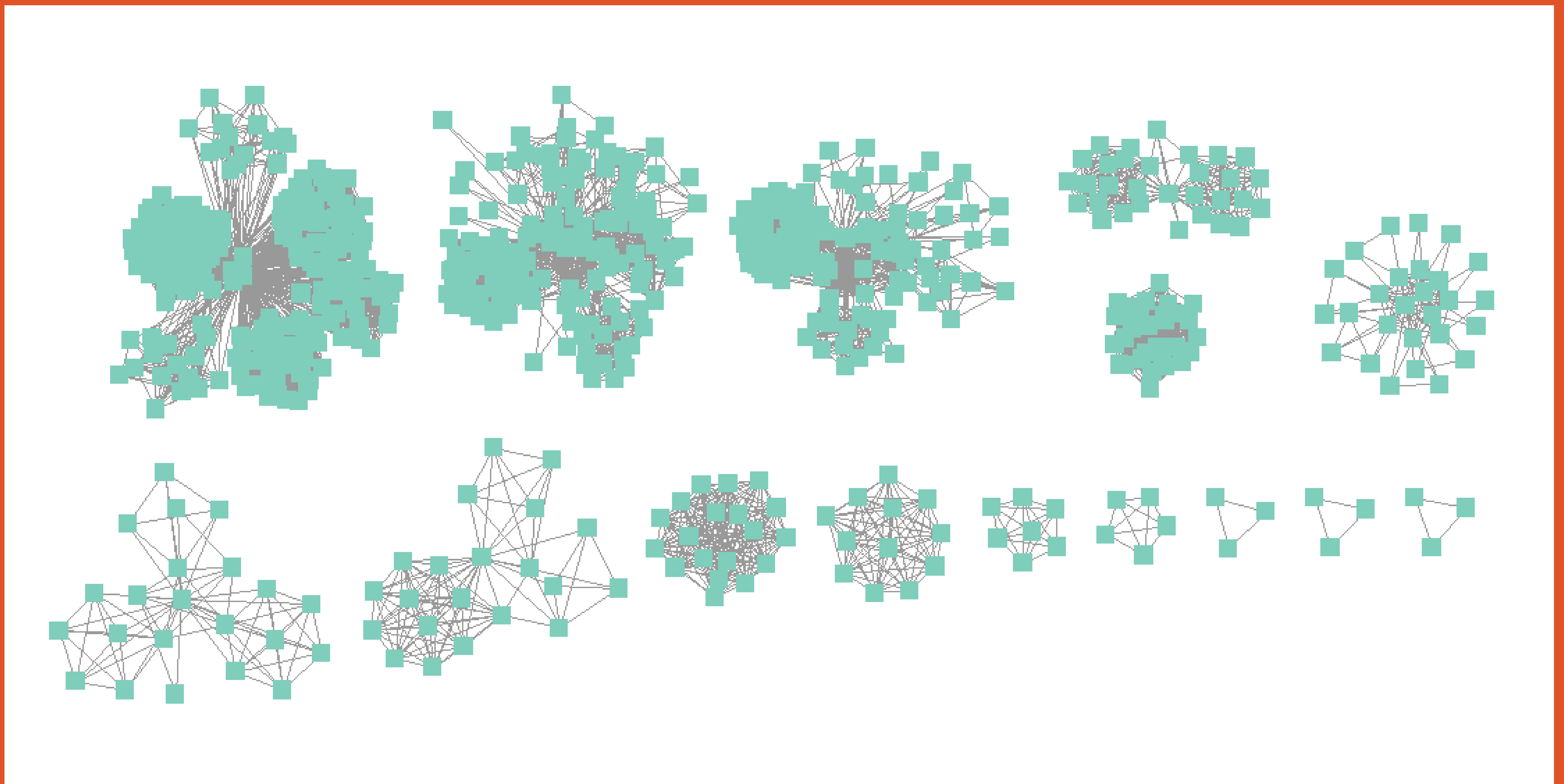
We begin to answer these questions studying the tripartite network for the Gymnosperm taxonomic group as a single network. The original network has a hierarchical structure that is difficult to gain information from, thus we will project this network. The three kinds of nodes for this tripartite network are the taxonomic family, genus, and species.

Method

We began with data from the *TRY Plant Trait Database*. There were 66,043 species from ~10 groups in the database. Our analysis was narrowed to one group, the Gymnosperms, plants with unprotected seeds including conifers, cycads, and ginkos. In this group, there are 15 families, 77 genera, and 647 species. Links were formed between species in the same genus, genera in the same family, species and its genera, genera and its family, and species and its family. Cytoscape was used to analyze this projection network.

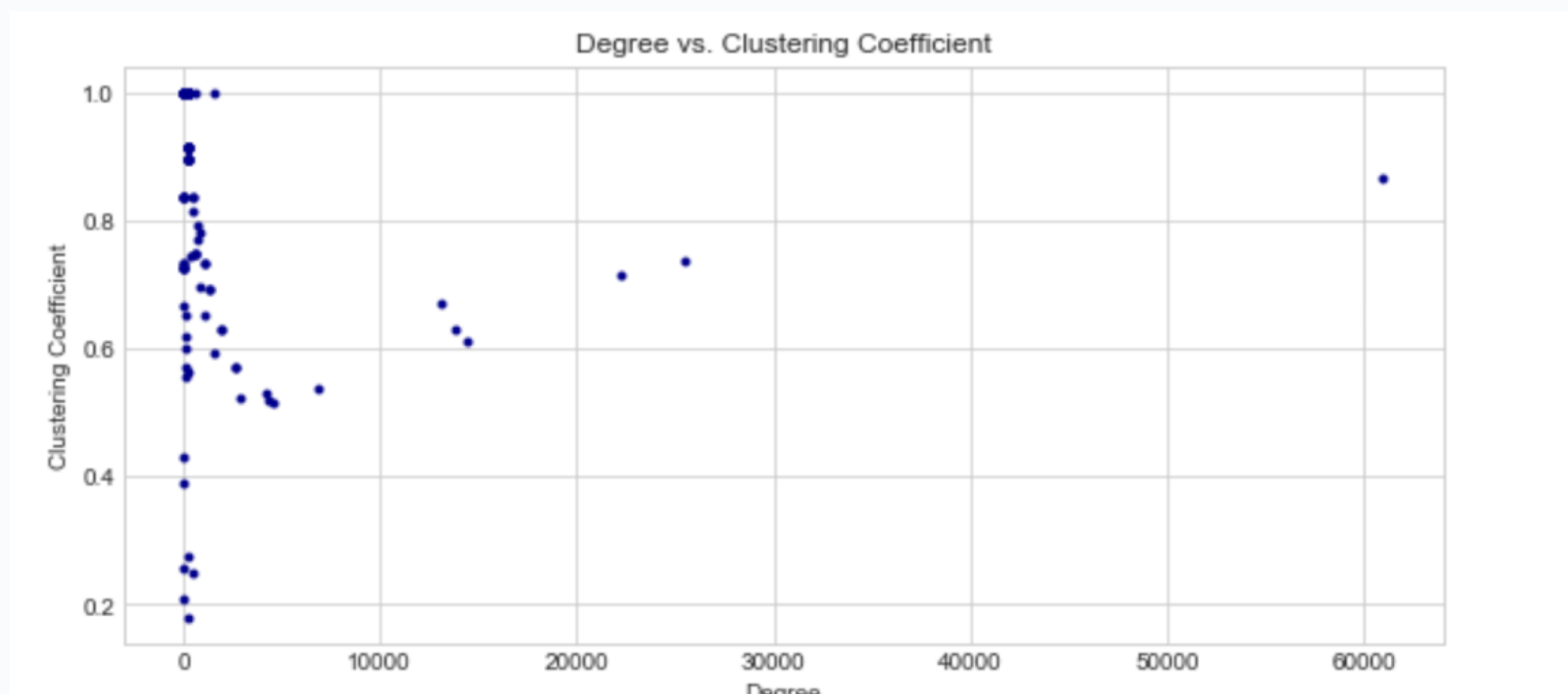
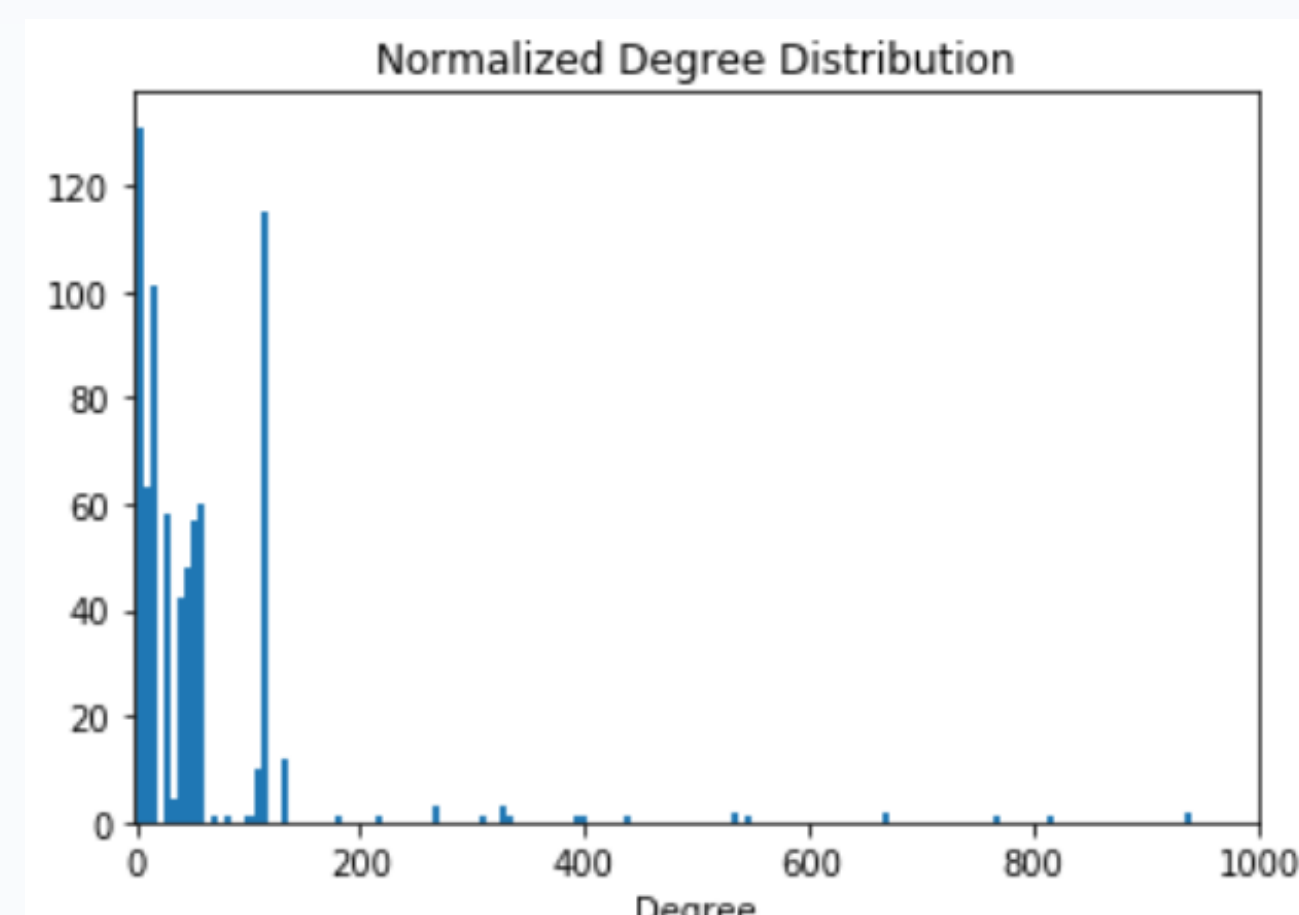


	D	E	F	G	H
1	Genus	Species/Ethet	Family	PhylogeneticGroup	PlantGrowthForm
2	Abrotia	maritima	Nyctaginaceae	Angiosperm_Eudicotyl	herb
3	Abutilon	theophrasti	Malvaceae	Angiosperm_Eudicotyl	herb
4	Acacia	anura	Fabaceae	Angiosperm_Eudicotyl	shrub/tree
5	Acacia	bivenata	Fabaceae	Angiosperm_Eudicotyl	shrub/tree
6	Acacia	busilini	Fabaceae	Angiosperm_Eudicotyl	shrub/tree
7	Acacia	cognata	Fabaceae	Angiosperm_Eudicotyl	shrub/tree
8	Acacia	deubata	Fabaceae	Angiosperm_Eudicotyl	shrub/tree
9	Acacia	elata	Fabaceae	Angiosperm_Eudicotyl	tree
10	Acacia	impinea	Fabaceae	Angiosperm_Eudicotyl	shrub/tree
11	Acacia	irrorata	Fabaceae	Angiosperm_Eudicotyl	shrub/tree
12	Acacia	jonesii	Fabaceae	Angiosperm_Eudicotyl	shrub
13	Acacia	koa	Fabaceae	Angiosperm_Eudicotyl	tree
14	Acacia	ligulata	Fabaceae	Angiosperm_Eudicotyl	shrub/tree
15	Acacia	mearnsii	Fabaceae	Angiosperm_Eudicotyl	tree
16	Acacia	melanoxylon	Fabaceae	Angiosperm_Eudicotyl	tree
17	Acacia	penulata	Fabaceae	Angiosperm_Eudicotyl	tree



Results

Number of Nodes: 739
 Number of Links: 131,321
 Connected Components: 15
 Clustering Coefficient: 0.986
 Smallest Degree: 2
 Lowest Degree: 60,950



Discussion

For our plant taxonomic network projection, the number of nodes is the sum of the families, genera, and species in the taxonomic group, Gymnosperms, thus is exactly what was expected. The number of possible links for 739 nodes follows $\frac{N(N-1)}{2} = 272,691$, whereas the number of links is 131,321. This value makes is what one might expect, since the nodes from different families (components) are by definition not able to form a link. There are 15 connected components, just what one might predict since there are 15 families in our dataset. The clustering coefficient is very close to 1. This is what one would expect for each component is very near fully connected. Since the network is unweighted, the degree distribution does not give us the most useful information.

Moving Forward

We should transform this unweighted, projected network to a weighted network. To as the original question, does the taxonomic classification correlate with genetic classifications, we need to compare the results of the analysis of the weighted network to a dataset with ground truth data on genetic characteristics.

Acknowledgments

The *TRY* initiative and data base is hosted, developed and maintained by J. Kattge and G. Bönisch (Max Planck Institute for Biogeochemistry, Jena, Germany).
 Cytoscape Program

Plant Taxonomic Network Projection

Chloe T. Calderon

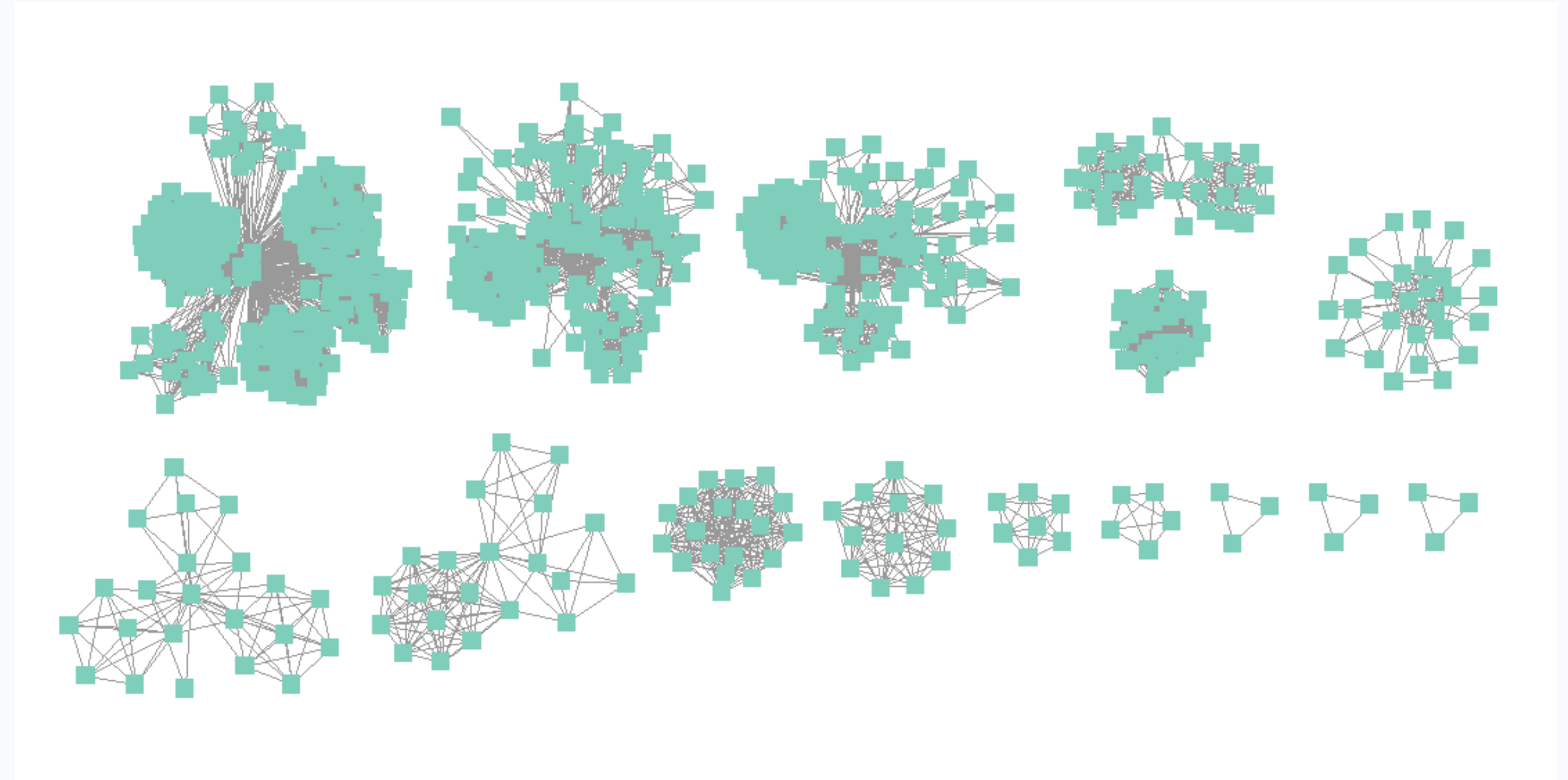
Department of Physics and Astronomy - Northwestern University

Abstract

Plant taxonomic classification is based on observable, macroscopic shared characteristics.

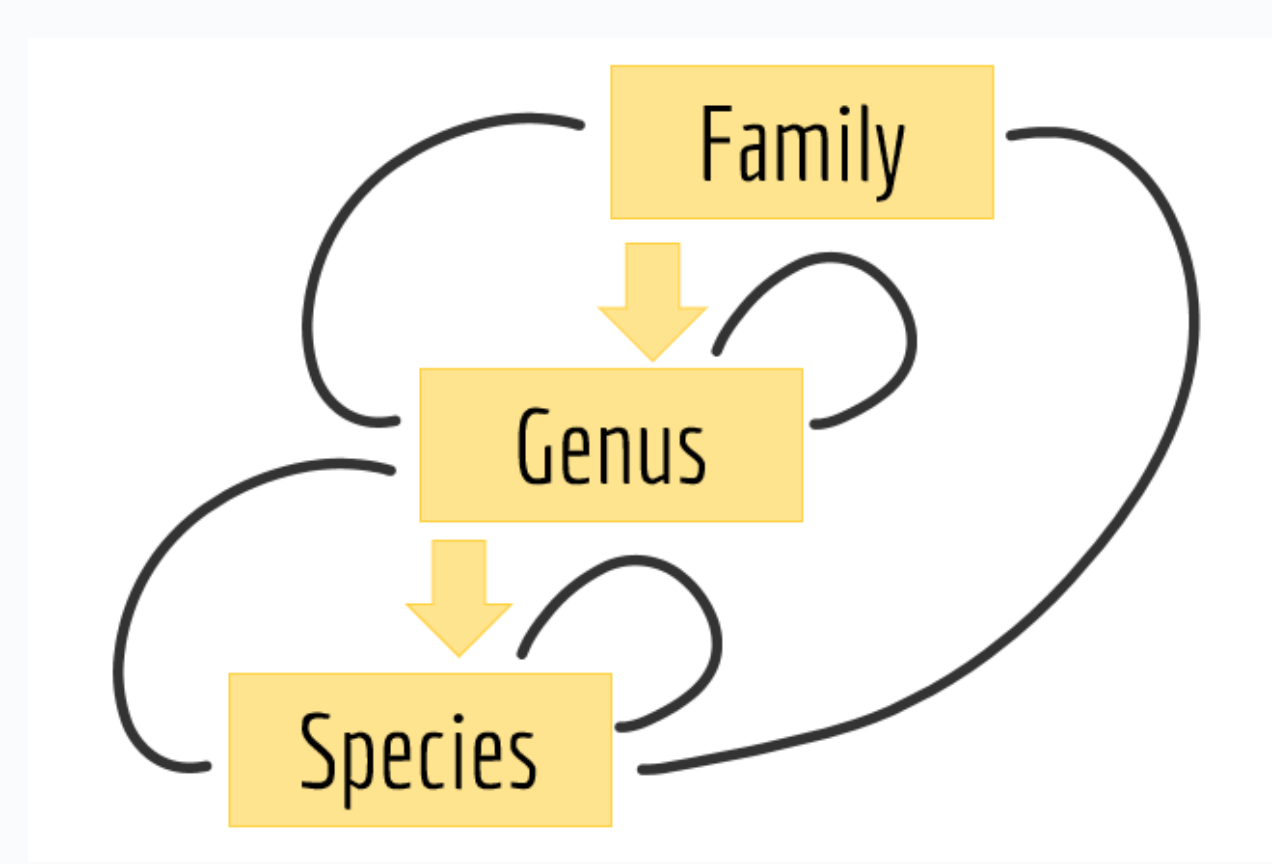
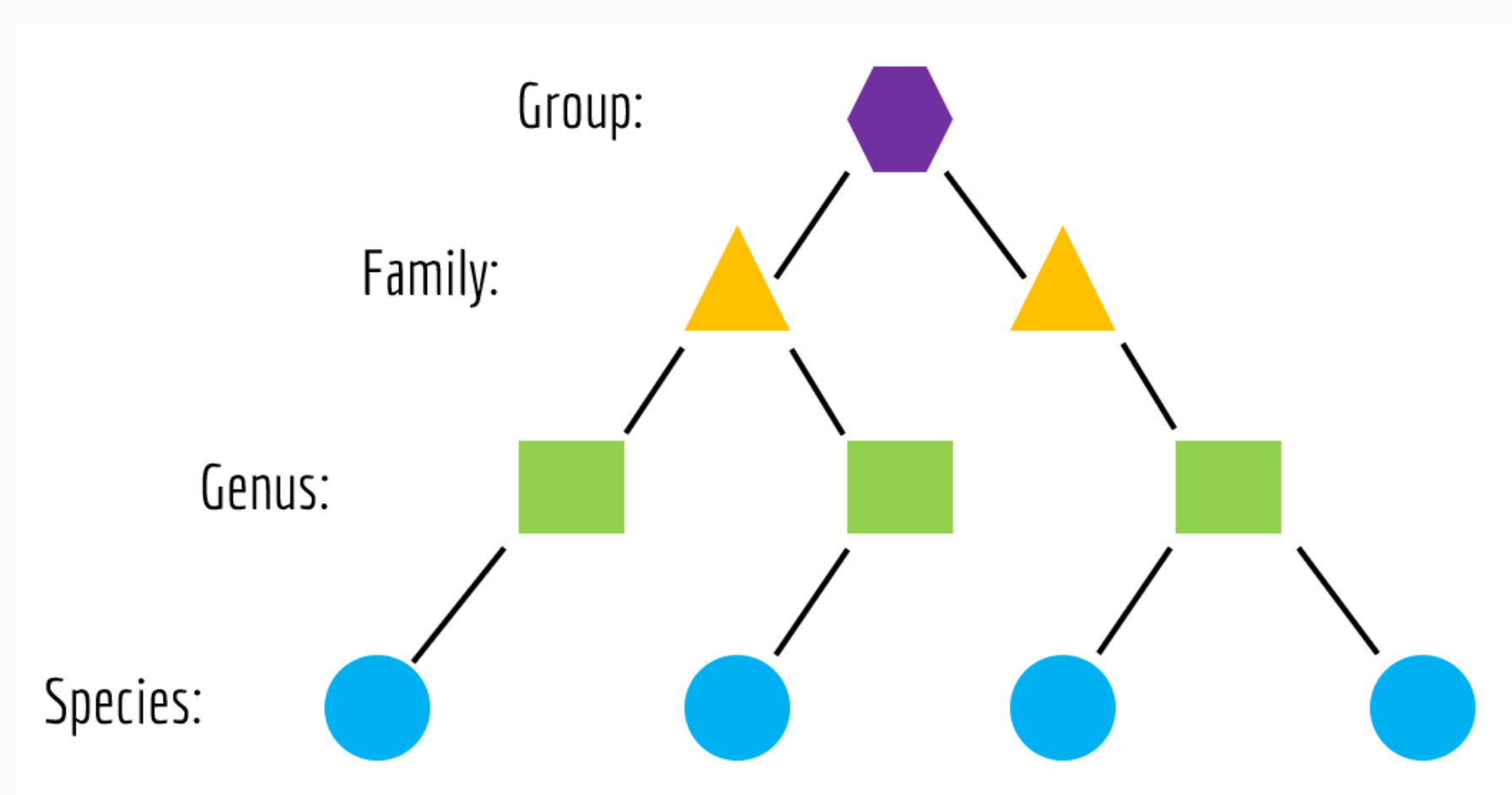
What would a classification system based on genetic characteristics (number of genes, number of copies, etc.) look like? How do genetic characteristics relate to our current classification system?

We begin to answer these questions studying the tripartite network for the Gymnosperm taxonomic group as a single network. The original network has a hierarchical structure that is difficult to gain information from, thus we will project this network. The three kinds of nodes for this tripartite network are the taxonomic family, genus, and species.



Method

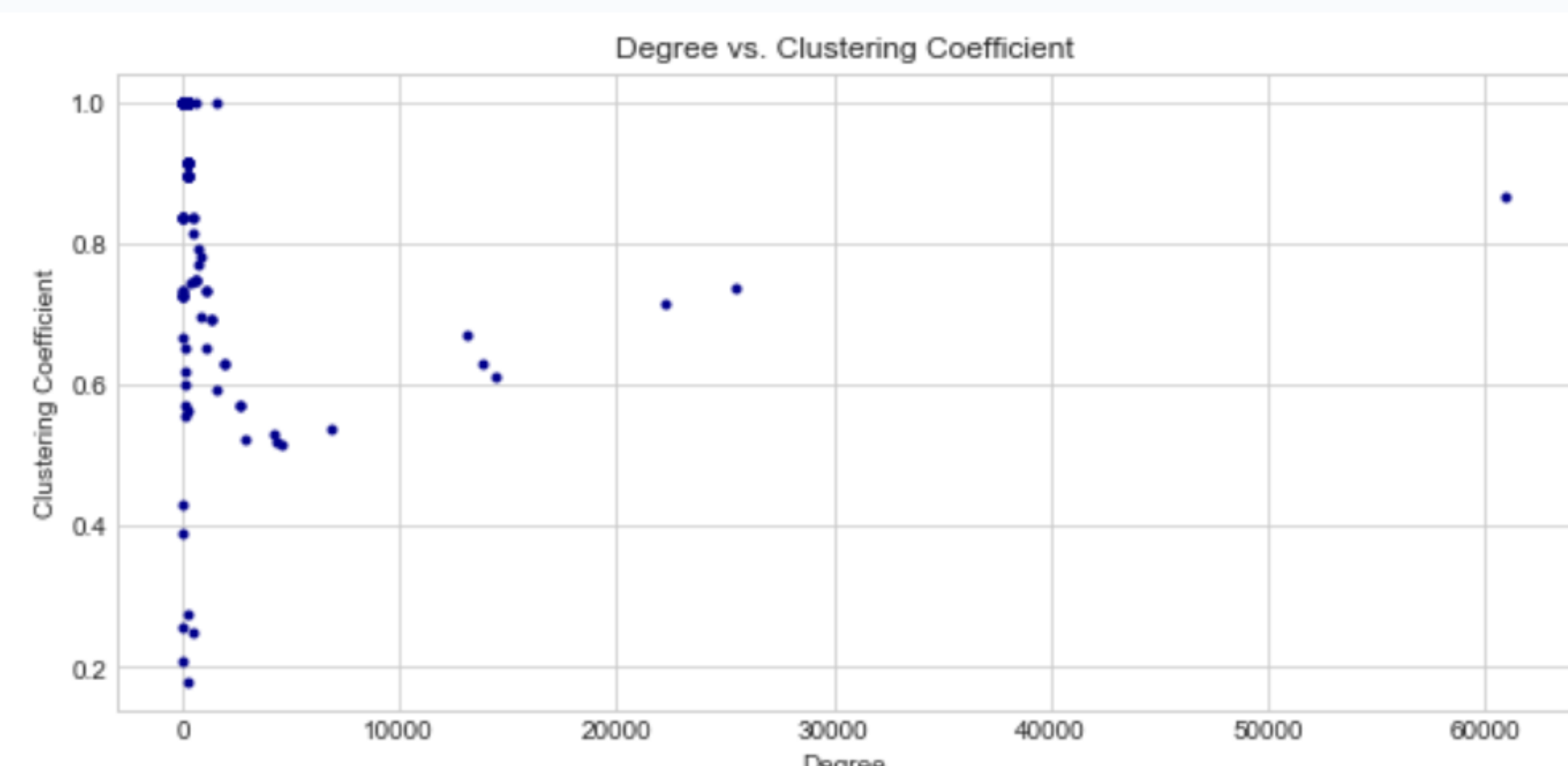
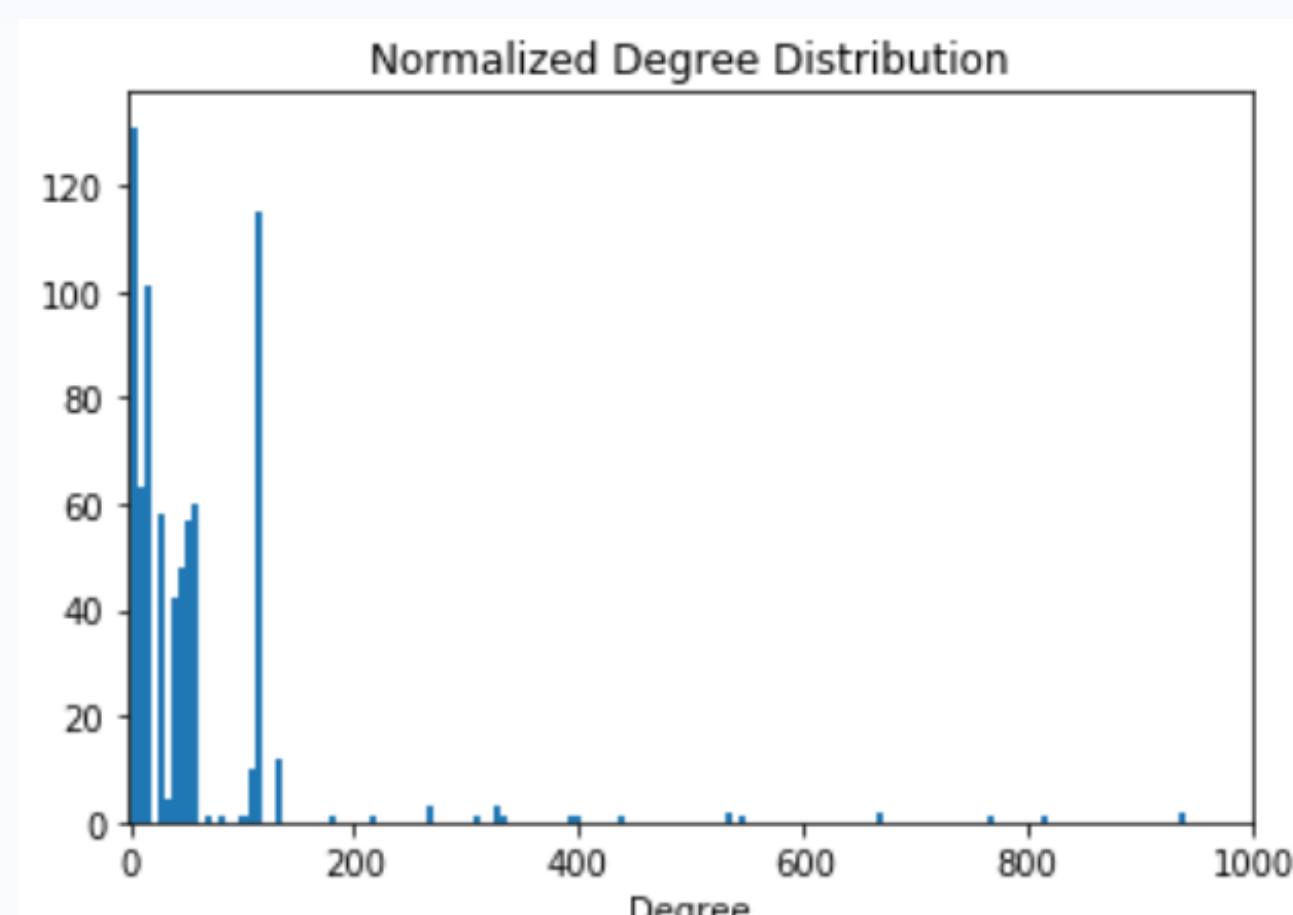
We began with data from the *TRY Plant Trait Database*. There were 66,043 species from ~10 groups in the database. Our analysis was narrowed to one group, the Gymnosperms, plants with unprotected seeds including conifers, cycads, and ginkos. In this group, there are 15 families, 77 genera, and 647 species. Links were formed between species in the same genus, genera in the same family, species and its genera, genera and its family, and species and its family. Cytoscape was used to analyze this projection network.



	D	E	F	G	H
1	Genus	SpeciesEpithet	Family	PhylogeneticGroup	PlantGrowthForm
2	Abronia	maritima	Nyctaginaceae	Angiosperm_Eudicotyl	herb
3	Abutilon	theophrasti	Malvaceae	Angiosperm_Eudicotyl	herb
4	Acacia	aneura	Fabaceae	Angiosperm_Eudicotyl	shrub/tree
5	Acacia	binervata	Fabaceae	Angiosperm_Eudicotyl	shrub/tree
6	Acacia	burkittii	Fabaceae	Angiosperm_Eudicotyl	shrub/tree
7	Acacia	cognata	Fabaceae	Angiosperm_Eudicotyl	shrub/tree
8	Acacia	dealbata	Fabaceae	Angiosperm_Eudicotyl	shrub/tree
9	Acacia	elata	Fabaceae	Angiosperm_Eudicotyl	tree
10	Acacia	implexa	Fabaceae	Angiosperm_Eudicotyl	shrub/tree
11	Acacia	irrorata	Fabaceae	Angiosperm_Eudicotyl	shrub/tree
12	Acacia	jonesii	Fabaceae	Angiosperm_Eudicotyl	shrub
13	Acacia	koa	Fabaceae	Angiosperm_Eudicotyl	tree
14	Acacia	ligulata	Fabaceae	Angiosperm_Eudicotyl	shrub/tree
15	Acacia	mearnsii	Fabaceae	Angiosperm_Eudicotyl	tree
16	Acacia	melanoxylon	Fabaceae	Angiosperm_Eudicotyl	tree
17	Acacia	pendula	Fabaceae	Angiosperm_Eudicotyl	tree

Results

Number of Nodes: 739
Number of Links: 131,321
Connected Components: 15
Clustering Coefficient: 0.986
Smallest Degree: 2
Lowest Degree: 60,950



Discussion

For our plant taxonomic network projection, the number of nodes is the sum of the families, genera, and species in the taxonomic group, Gymnosperms, thus is exactly what was expected. The number of possible links for 739 nodes follows $\frac{N(N-1)}{2} = 272,691$, whereas the number of links is 131,321. This value makes is what one might expect, since the nodes from different families (components) are by definition not able to form a link. There are 15 connected components, just what one might predict since there are 15 families in our dataset. The clustering coefficient is very close to 1. This is what one would expect for each component is very near fully connected. Since the network is unweighted, the degree distribution does not give us the most useful information.

Moving Forward

We should transform this unweighted, projected network to a weighted network. To as the original question, does the taxonomic classification correlate with genetic classifications, we need to compare the results of the analysis of the weighted network to a dataset with ground truth data on genetic characteristics.

Acknowledgments

The *TRY* initiative and data base is hosted, developed and maintained by J. Kattge and G. Bönisch (Max Planck Institute for Biogeochemistry, Jena, Germany).
Cytoscape Program

Plant Taxonomic Network Projection

Chloe T. Calderon

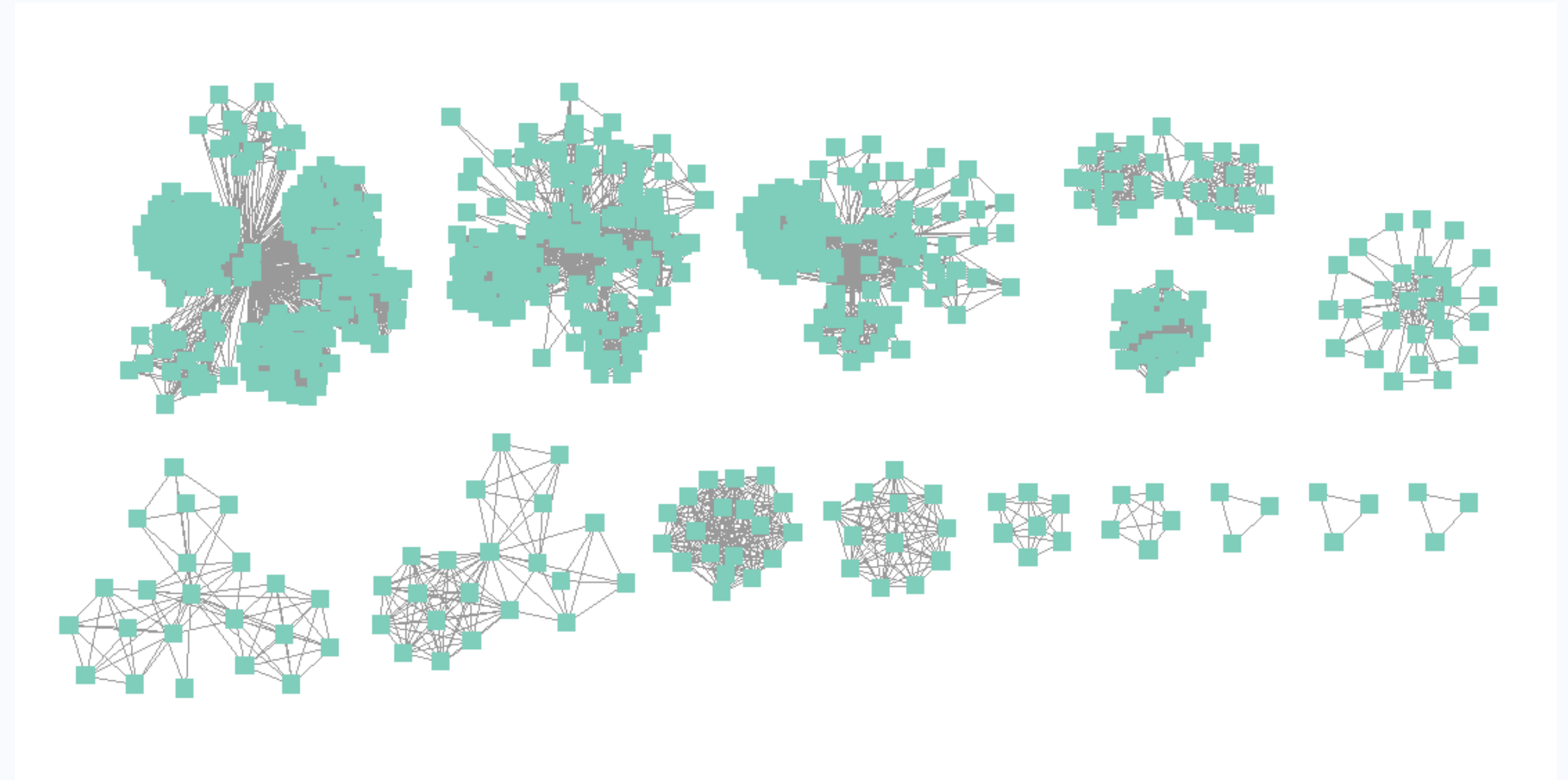
Department of Physics and Astronomy - Northwestern University

Abstract

Plant taxonomic classification is based on observable, macroscopic shared characteristics.

What would a classification system based on genetic characteristics (number of genes, number of copies, etc.) look like? How do genetic characteristics relate to our current classification system?

We begin to answer these questions studying the tripartite network for the Gymnosperm taxonomic group as a single network. The original network has a hierarchical structure that is difficult to gain information from, thus we will project this network. The three kinds of nodes for this tripartite network are the taxonomic family, genus, and species.

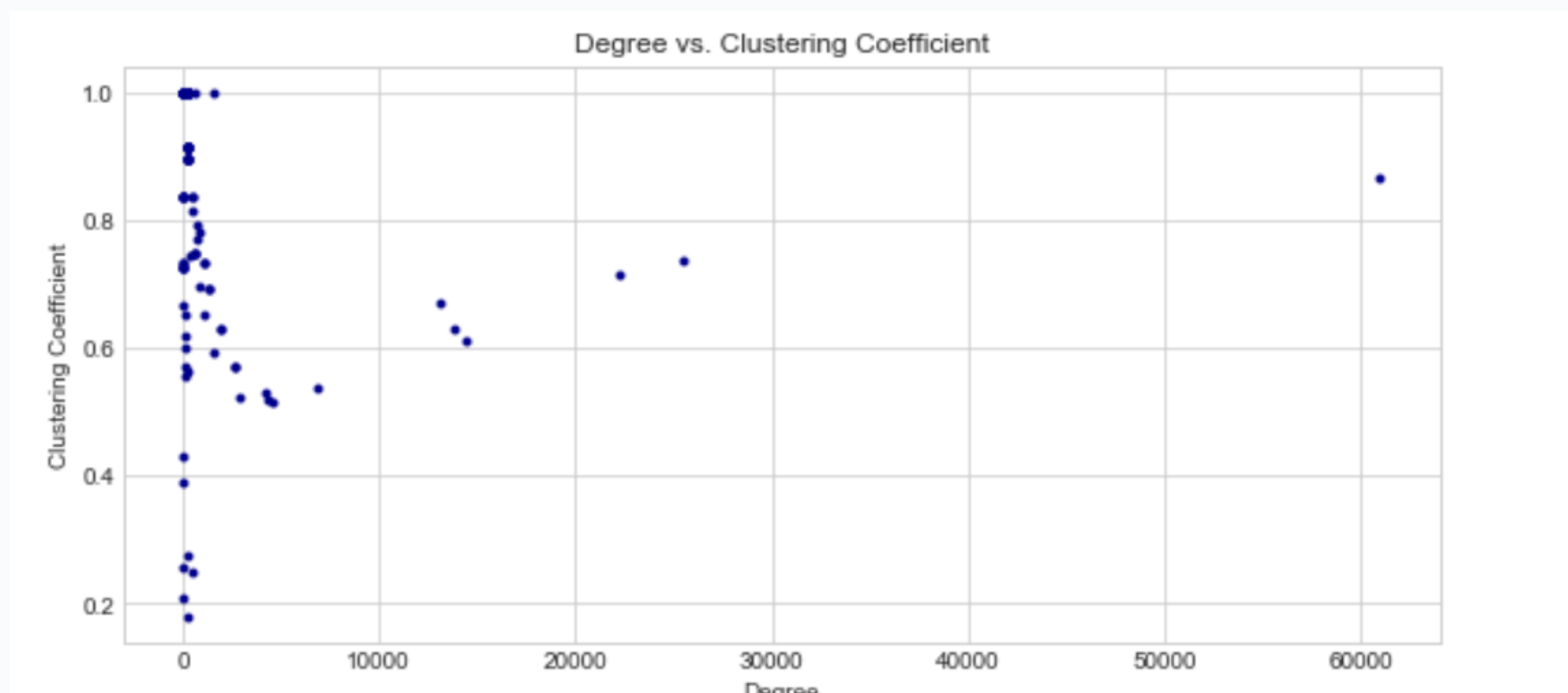
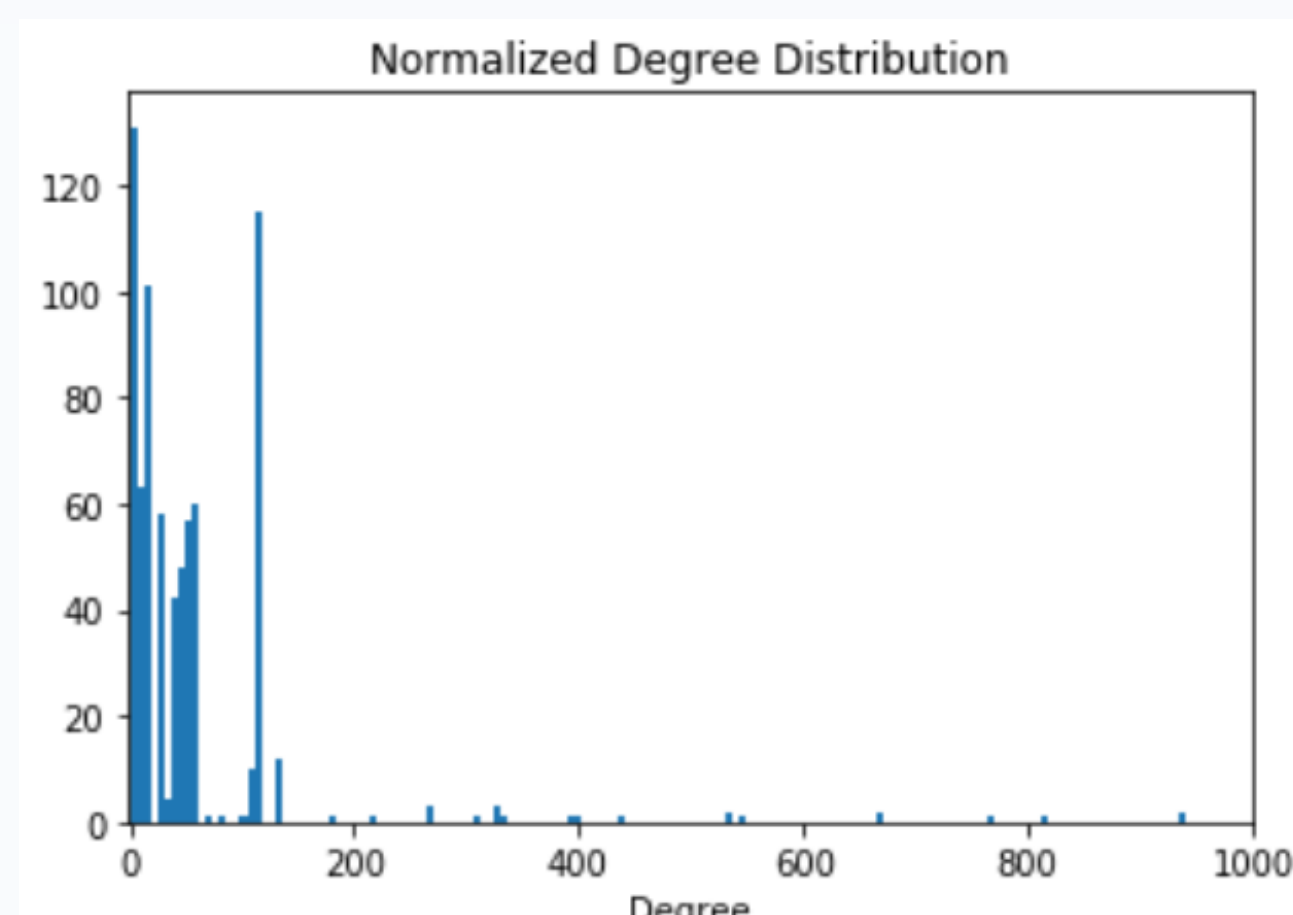


Method

	D	E	F	G	H
1	Genus	SpeciesEpithet	Family	PhylogeneticGroup	PlantGrowthForm
2	Abrotonia	maritima	Nyctaginaceae	Angiosperm_Eudicotyl	herb
3	Abutilon	theophrasti	Malvaceae	Angiosperm_Eudicotyl	herb
4	Acacia	aneura	Fabaceae	Angiosperm_Eudicotyl	shrub/tree
5	Acacia	binervata	Fabaceae	Angiosperm_Eudicotyl	shrub/tree
6	Acacia	burkittii	Fabaceae	Angiosperm_Eudicotyl	shrub/tree
7	Acacia	cognata	Fabaceae	Angiosperm_Eudicotyl	shrub/tree
8	Acacia	dealbata	Fabaceae	Angiosperm_Eudicotyl	shrub/tree
9	Acacia	elata	Fabaceae	Angiosperm_Eudicotyl	tree
10	Acacia	implexa	Fabaceae	Angiosperm_Eudicotyl	shrub/tree
11	Acacia	irrorata	Fabaceae	Angiosperm_Eudicotyl	shrub/tree
12	Acacia	jonesii	Fabaceae	Angiosperm_Eudicotyl	shrub
13	Acacia	koa	Fabaceae	Angiosperm_Eudicotyl	tree
14	Acacia	ligulata	Fabaceae	Angiosperm_Eudicotyl	shrub/tree
15	Acacia	mearnsii	Fabaceae	Angiosperm_Eudicotyl	tree
16	Acacia	melanoxyton	Fabaceae	Angiosperm_Eudicotyl	tree
17	Acacia	pendula	Fabaceae	Angiosperm_Eudicotyl	tree

Results

Number of Nodes: 739
Number of Links: 131,321
Connected Components: 15
Clustering Coefficient: 0.986
Smallest Degree: 2
Lowest Degree: 60,950



Discussion

For our plant taxonomic network projection, the number of nodes is the sum of the families, genera, and species in the taxonomic group, Gymnosperms, thus is exactly what was expected. The number of possible links for 739 nodes follows $\frac{N(N-1)}{2} = 272,691$, whereas the number of links is 131,321. This value makes is what one might expect, since the nodes from different families (components) are by definition not able to form a link. There are 15 connected components, just what one might predict since there are 15 families in our dataset. The clustering coefficient is very close to 1. This is what one would expect for each component is very near fully connected. Since the network is unweighted, the degree distribution does not give us the most useful information.

Moving Forward

We should transform this unweighted, projected network to a weighted network. To as the original question, does the taxonomic classification correlate with genetic classifications, we need to compare the results of the analysis of the weighted network to a dataset with ground truth data on genetic characteristics.

Acknowledgments

The TRY initiative and data base is hosted, developed and maintained by J. Kattge and G. Bönisch (Max Planck Institute for Biogeochemistry, Jena, Germany).
Cytoscape Program

Plant Taxonomic Network Projection

Chloe T. Calderon

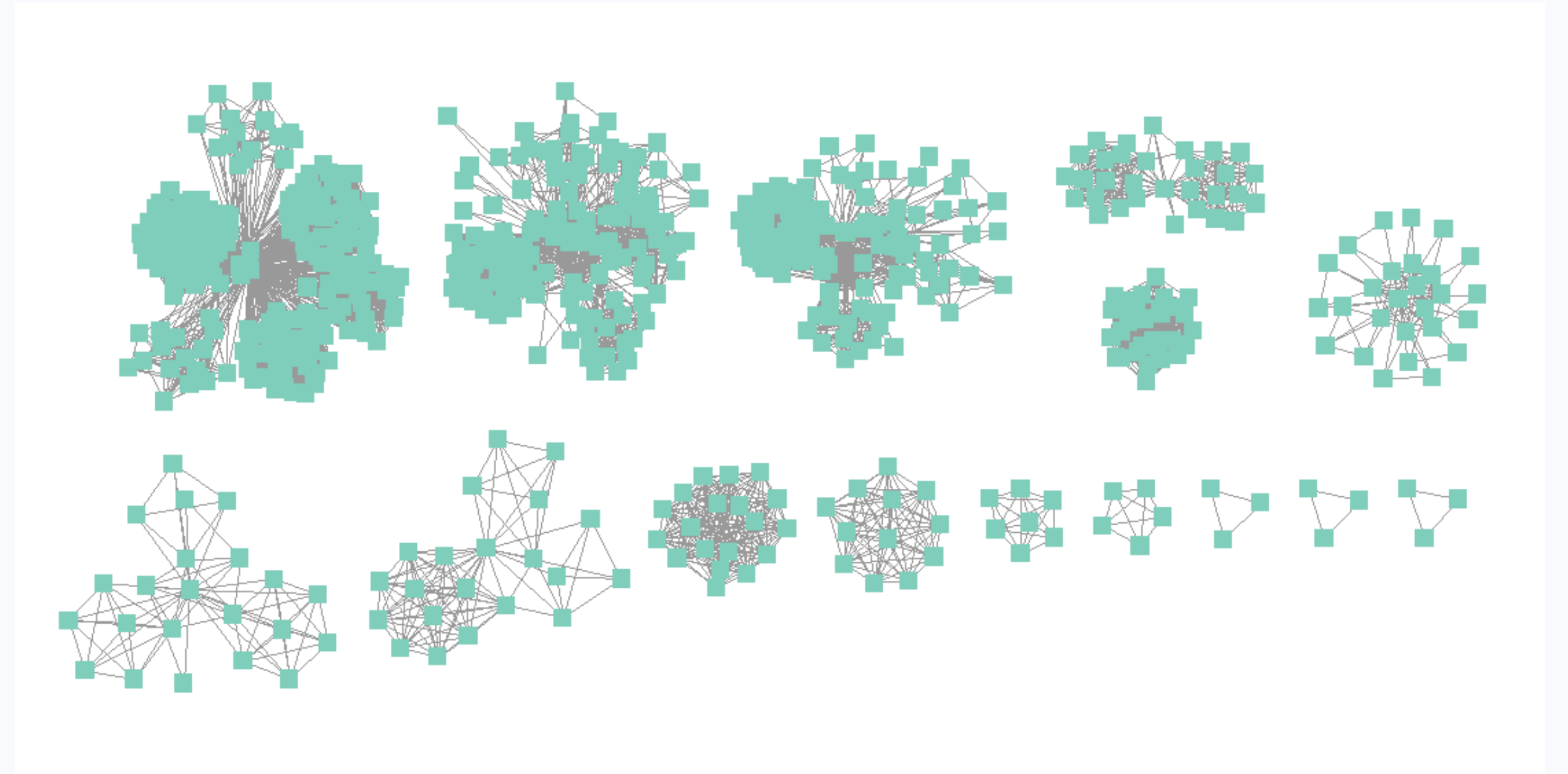
Department of Physics and Astronomy - Northwestern University

Abstract

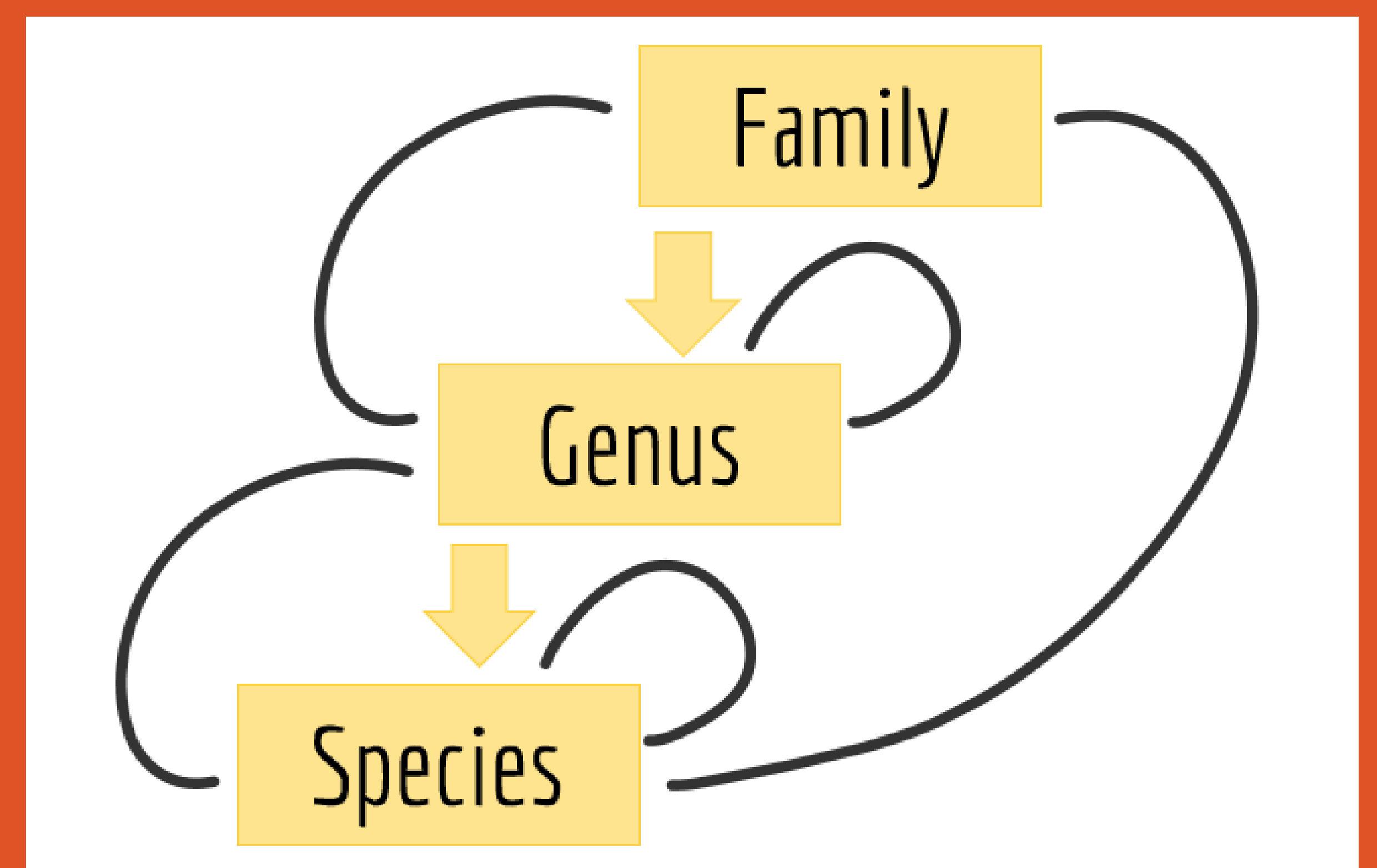
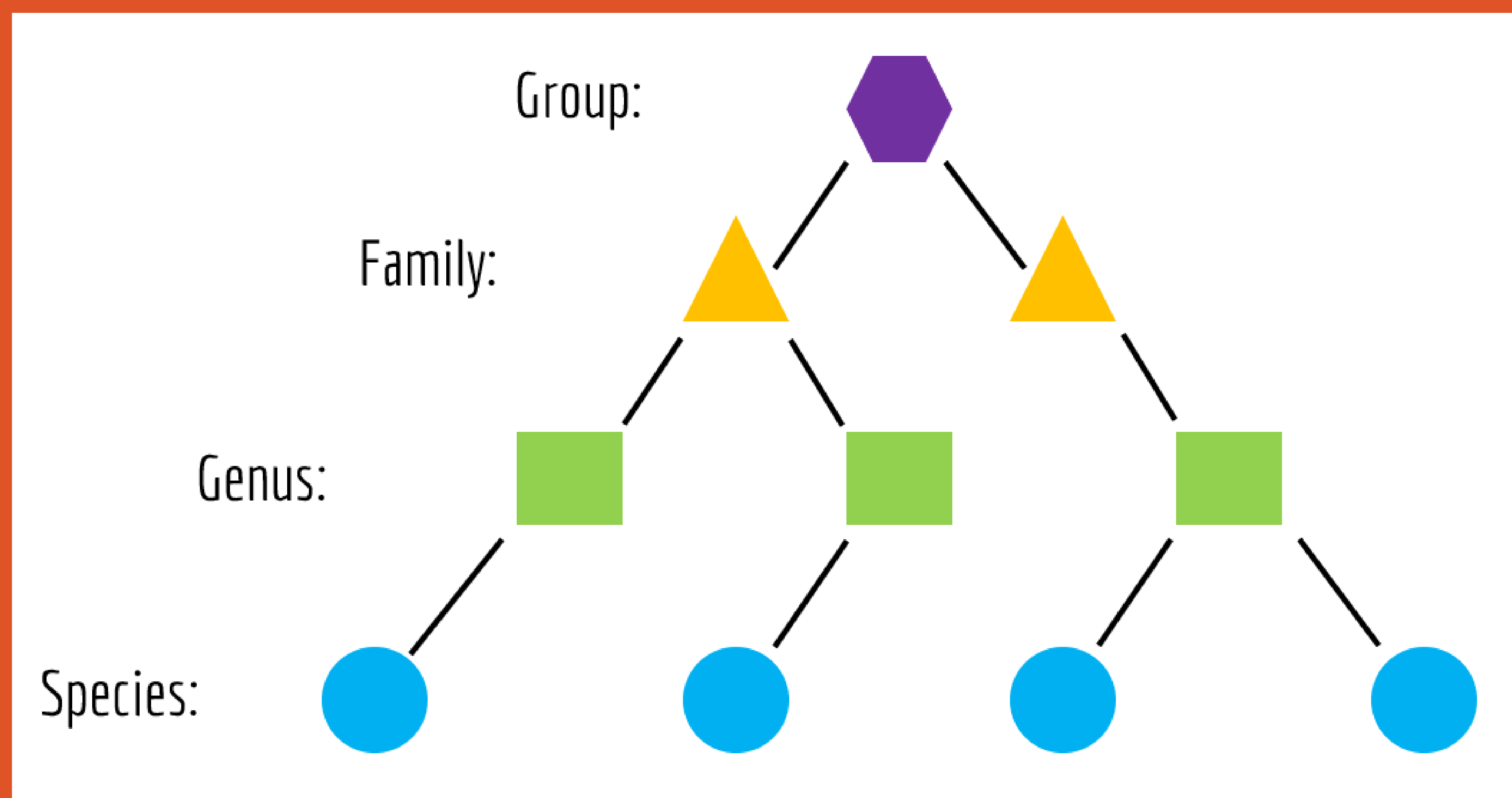
Plant taxonomic classification is based on observable, macroscopic shared characteristics.

What would a classification system based on genetic characteristics (number of genes, number of copies, etc.) look like? How do genetic characteristics relate to our current classification system?

We begin to answer these questions studying the tripartite network for the Gymnosperm taxonomic group as a single network. The original network has a hierarchical structure that is difficult to gain information from, thus we will project this network. The three kinds of nodes for this tripartite network are the taxonomic family, genus, and species.

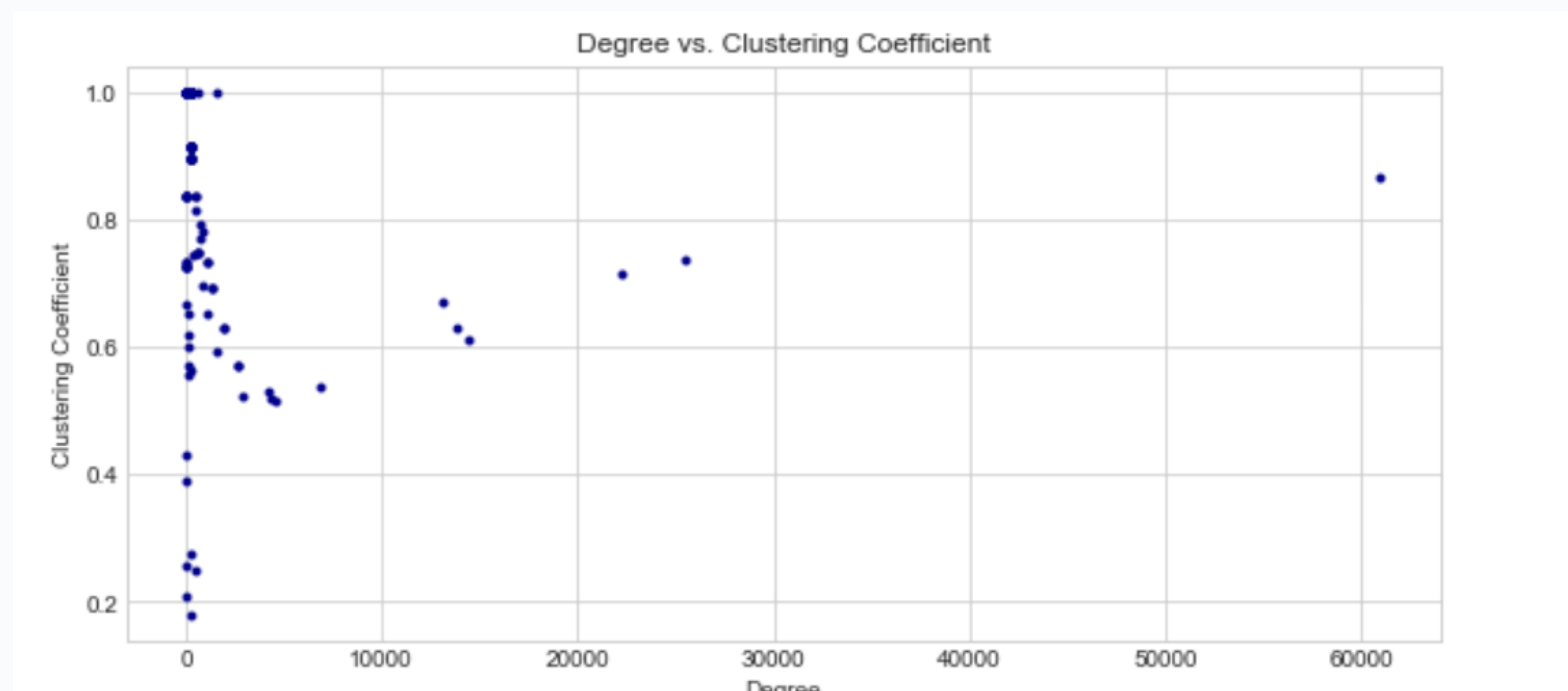
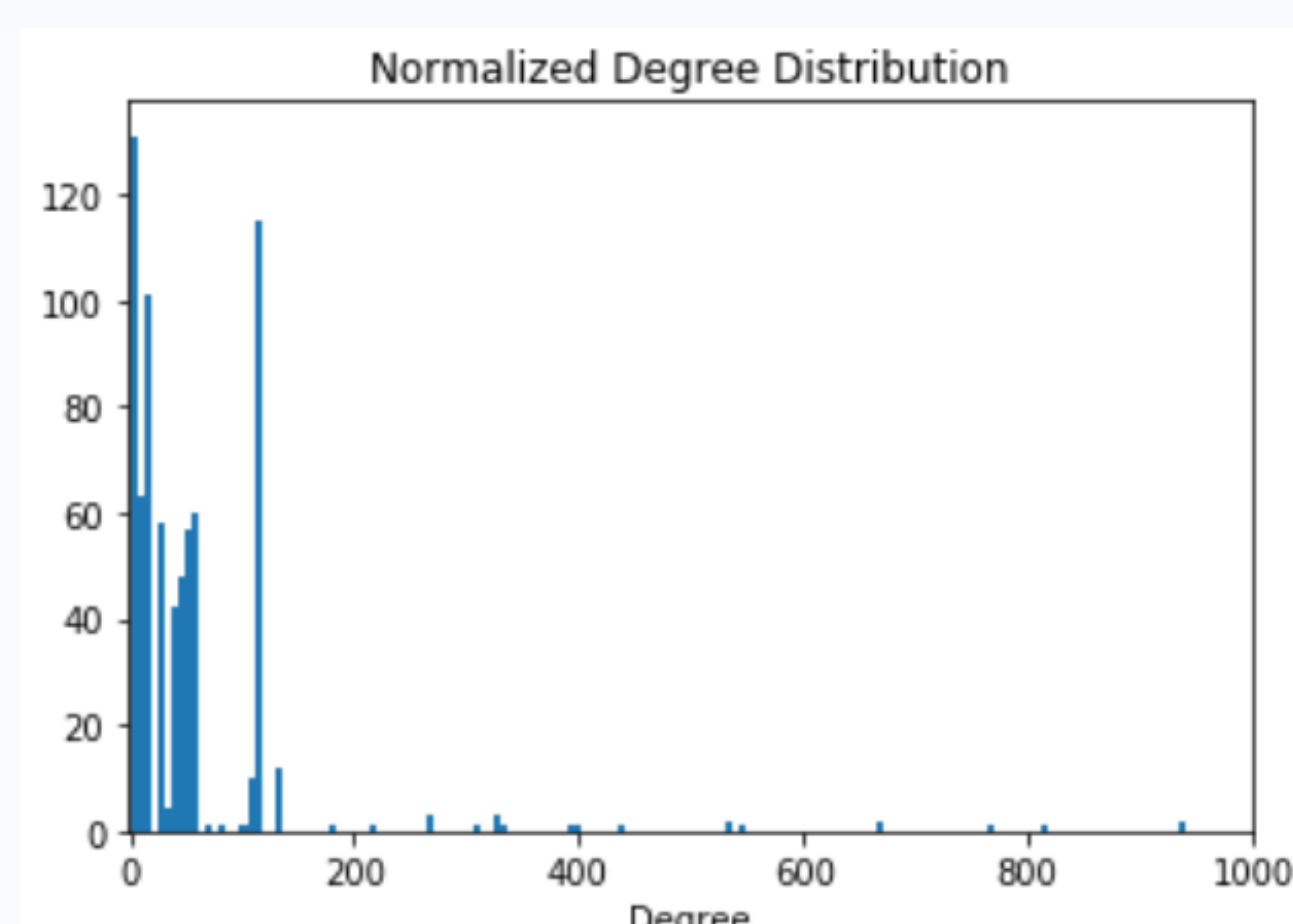


Method



Results

Number of Nodes: 739
Number of Links: 131,321
Connected Components: 15
Clustering Coefficient: 0.986
Smallest Degree: 2
Lowest Degree: 60,950



Discussion

For our plant taxonomic network projection, the number of nodes is the sum of the families, genera, and species in the taxonomic group, Gymnosperms, thus is exactly what was expected. The number of possible links for 739 nodes follows $\frac{N(N-1)}{2} = 272,691$, whereas the number of links is 131,321. This value makes is what one might expect, since the nodes from different families (components) are by definition not able to form a link. There are 15 connected components, just what one might predict since there are 15 families in our dataset. The clustering coefficient is very close to 1. This is what one would expect for each component is very near fully connected. Since the network is unweighted, the degree distribution does not give us the most useful information.

Moving Forward

We should transform this unweighted, projected network to a weighted network. To as the original question, does the taxonomic classification correlate with genetic classifications, we need to compare the results of the analysis of the weighted network to a dataset with ground truth data on genetic characteristics.

Acknowledgments

The *TRY* initiative and data base is hosted, developed and maintained by J. Kattge and G. Bönisch (Max Planck Institute for Biogeochemistry, Jena, Germany).
Cytoscape Program

Plant Taxonomic Network Projection

Chloe T. Calderon

Department of Physics and Astronomy - Northwestern University

Abstract

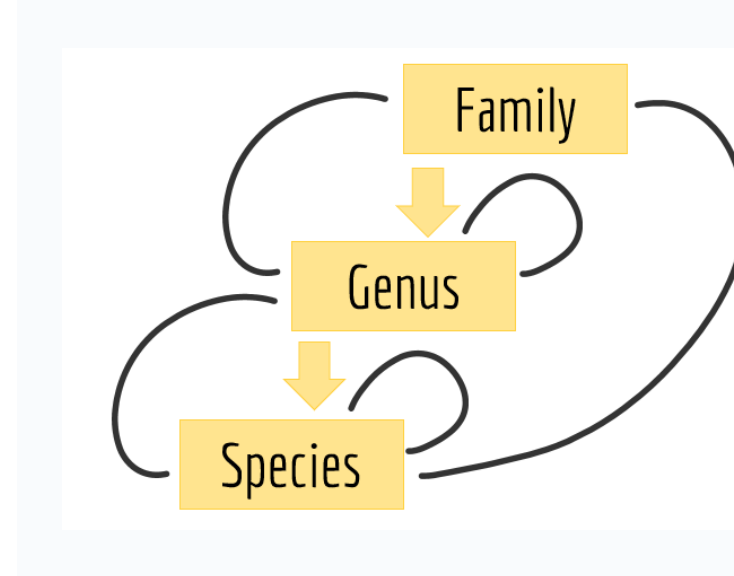
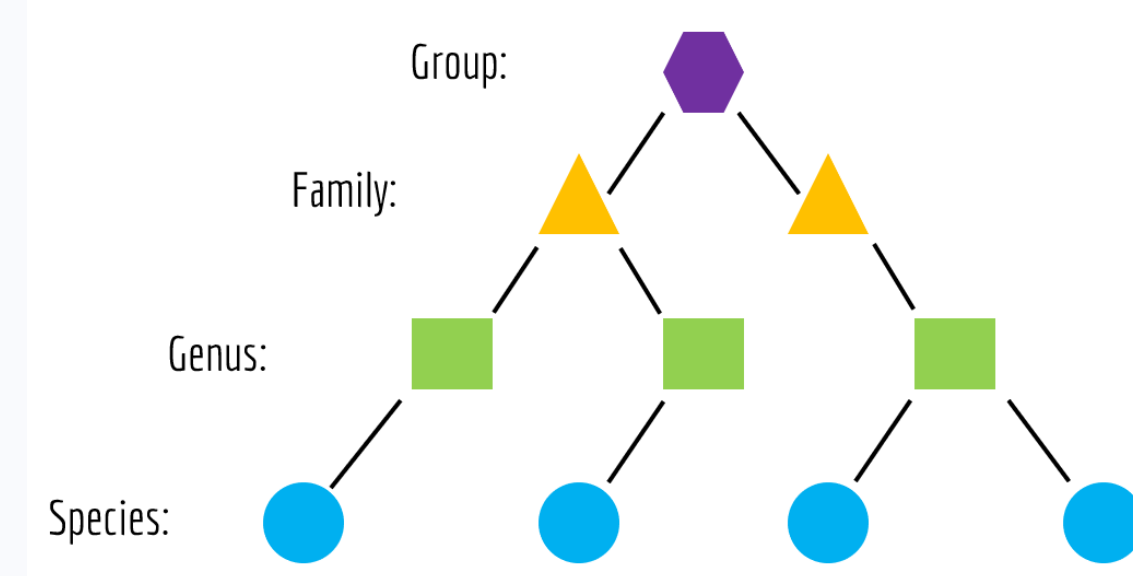
Plant taxonomic classification is based on observable, macroscopic shared characteristics.

What would a classification system based on genetic characteristics (number of genes, number of copies, etc.) look like? How do genetic characteristics relate to our current classification system?

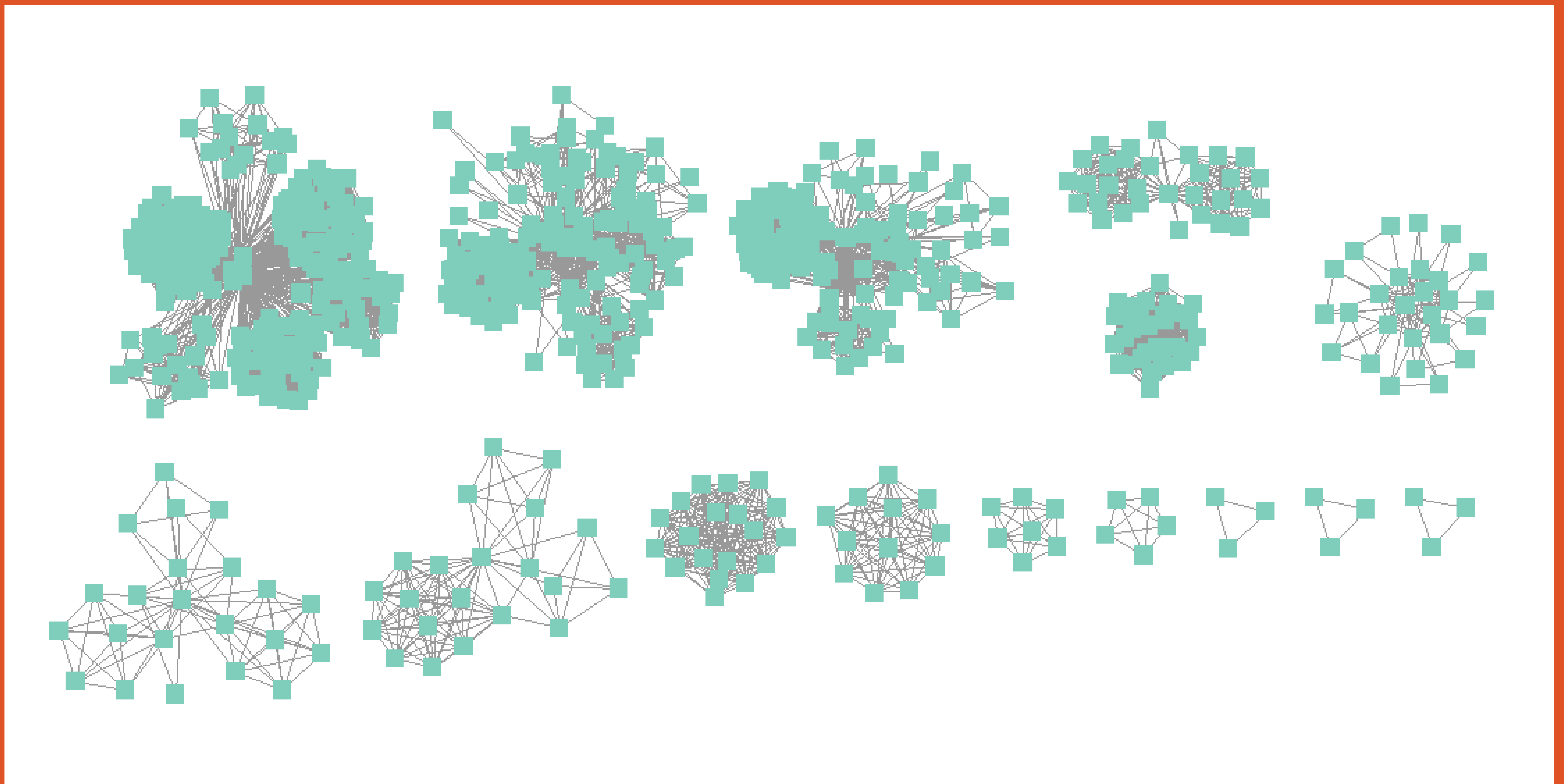
We begin to answer these questions studying the tripartite network for the Gymnosperm taxonomic group as a single network. The original network has a hierarchical structure that is difficult to gain information from, thus we will project this network. The three kinds of nodes for this tripartite network are the taxonomic family, genus, and species.

Method

We began with data from the *TRY Plant Trait Database*. There were 66,043 species from ~10 groups in the database. Our analysis was narrowed to one group, the Gymnosperms, plants with unprotected seeds including conifers, cycads, and ginkos. In this group, there are 15 families, 77 genera, and 647 species. Links were formed between species in the same genus, genera in the same family, species and its genera, genera and its family, and species and its family. Cytoscape was used to analyze this projection network.

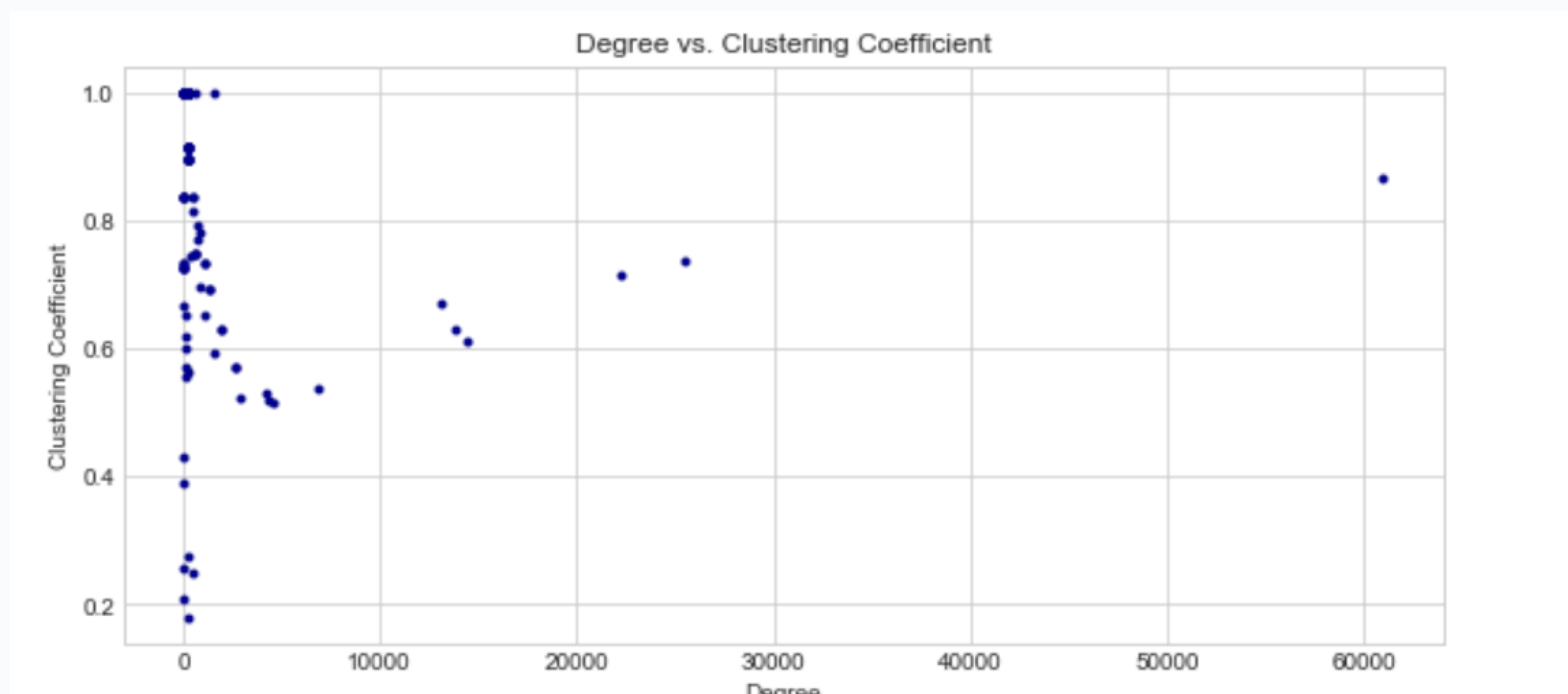
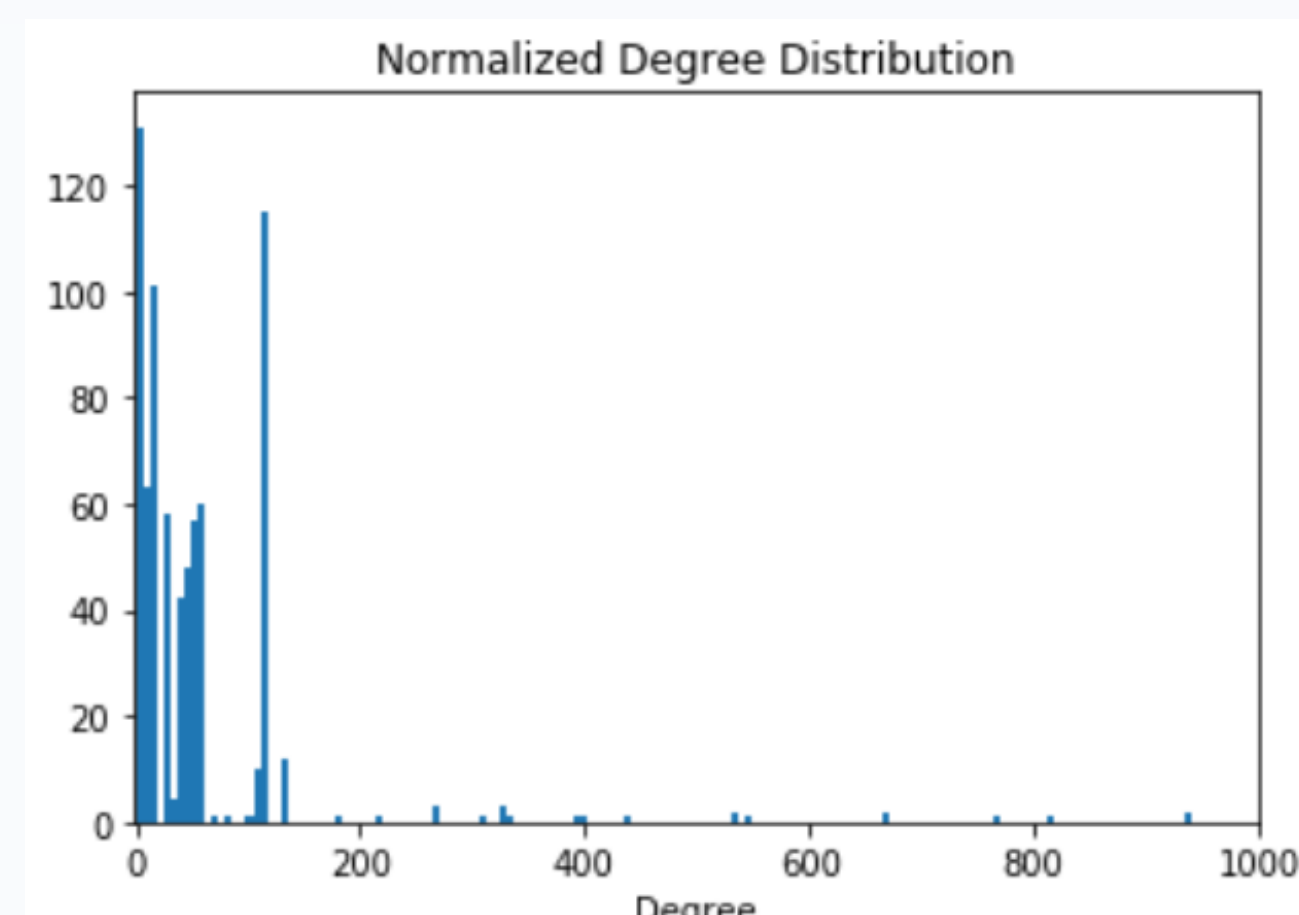


	D	E	F	G	H
1	Genus	Species/Ethet	Family	PhylogeneticGroup	PlantGrowthForm
2	Abrotia	maritima	Nyctaginaceae	Angiosperm_Eudicotyl	herb
3	Abutilon	theophrasti	Malvaceae	Angiosperm_Eudicotyl	herb
4	Acacia	anura	Fabaceae	Angiosperm_Eudicotyl	shrub/tree
5	Acacia	bivennata	Fabaceae	Angiosperm_Eudicotyl	shrub/tree
6	Acacia	buschii	Fabaceae	Angiosperm_Eudicotyl	shrub/tree
7	Acacia	cognata	Fabaceae	Angiosperm_Eudicotyl	shrub/tree
8	Acacia	deubata	Fabaceae	Angiosperm_Eudicotyl	shrub/tree
9	Acacia	elata	Fabaceae	Angiosperm_Eudicotyl	tree
10	Acacia	impinea	Fabaceae	Angiosperm_Eudicotyl	shrub/tree
11	Acacia	irrorata	Fabaceae	Angiosperm_Eudicotyl	shrub/tree
12	Acacia	jonesii	Fabaceae	Angiosperm_Eudicotyl	shrub
13	Acacia	koa	Fabaceae	Angiosperm_Eudicotyl	tree
14	Acacia	ligulata	Fabaceae	Angiosperm_Eudicotyl	shrub/tree
15	Acacia	mearnsii	Fabaceae	Angiosperm_Eudicotyl	tree
16	Acacia	melanoxylon	Fabaceae	Angiosperm_Eudicotyl	tree
17	Acacia	penulata	Fabaceae	Angiosperm_Eudicotyl	tree



Results

Number of Nodes: 739
 Number of Links: 131,321
 Connected Components: 15
 Clustering Coefficient: 0.986
 Smallest Degree: 2
 Lowest Degree: 60,950



Discussion

For our plant taxonomic network projection, the number of nodes is the sum of the families, genera, and species in the taxonomic group, Gymnosperms, thus is exactly what was expected. The number of possible links for 739 nodes follows $\frac{N(N-1)}{2} = 272,691$, whereas the number of links is 131,321. This value makes is what one might expect, since the nodes from different families (components) are by definition not able to form a link. There are 15 connected components, just what one might predict since there are 15 families in our dataset. The clustering coefficient is very close to 1. This is what one would expect for each component is very near fully connected. Since the network is unweighted, the degree distribution does not give us the most useful information.

Moving Forward

We should transform this unweighted, projected network to a weighted network. To as the original question, does the taxonomic classification correlate with genetic classifications, we need to compare the results of the analysis of the weighted network to a dataset with ground truth data on genetic characteristics.

Acknowledgments

The *TRY* initiative and data base is hosted, developed and maintained by J. Kattge and G. Bönisch (Max Planck Institute for Biogeochemistry, Jena, Germany).
 Cytoscape Program

Plant Taxonomic Network Projection

Chloe T. Calderon

Department of Physics and Astronomy - Northwestern University

Abstract

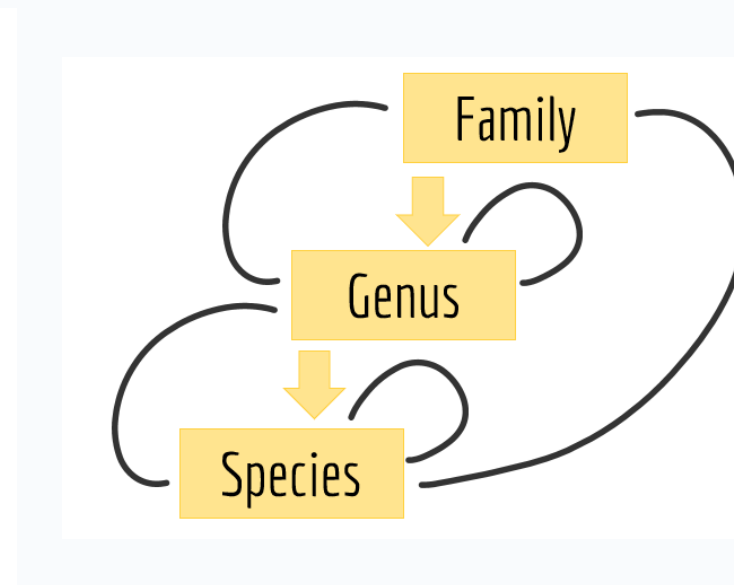
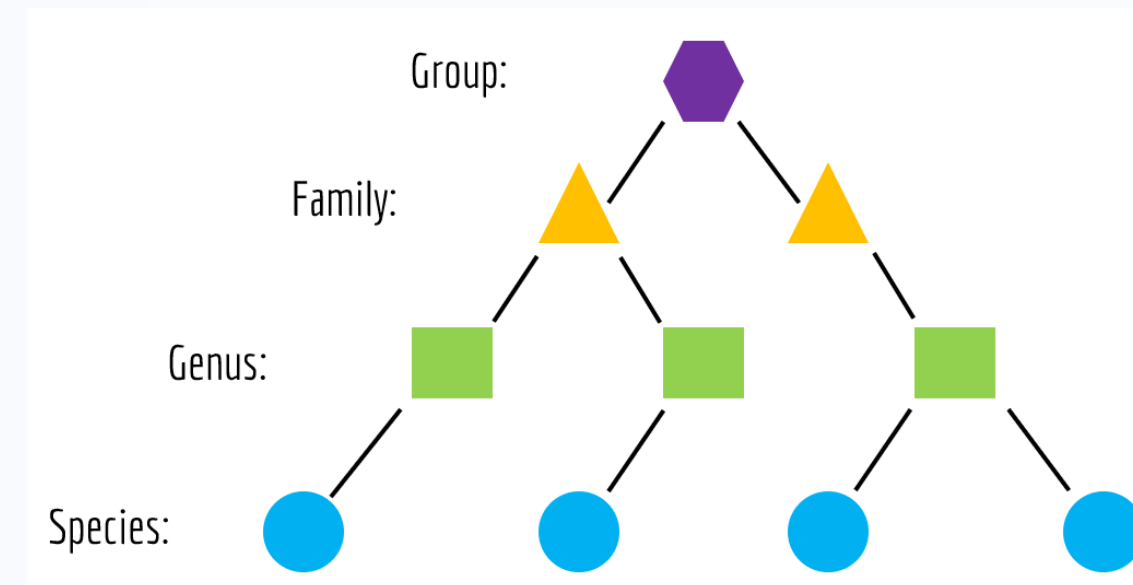
Plant taxonomic classification is based on observable, macroscopic shared characteristics.

What would a classification system based on genetic characteristics (number of genes, number of copies, etc.) look like? How do genetic characteristics relate to our current classification system?

We begin to answer these questions studying the tripartite network for the Gymnosperm taxonomic group as a single network. The original network has a hierarchical structure that is difficult to gain information from, thus we will project this network. The three kinds of nodes for this tripartite network are the taxonomic family, genus, and species.

Method

We began with data from the *TRY Plant Trait Database*. There were 66,043 species from ~10 groups in the database. Our analysis was narrowed to one group, the Gymnosperms, plants with unprotected seeds including conifers, cycads, and ginkos. In this group, there are 15 families, 77 genera, and 647 species. Links were formed between species in the same genus, genera in the same family, and species and its family. Cytoscape was used to analyze this projection network.



Genus	Species/Ethet	Family	PhylogeneticGroup	PlantGrowthForm
Abrotia	maritima	Nyctaginaceae	Angiosperm_Eudicotyl	herb
Abutilon	theophrasti	Malvaceae	Angiosperm_Eudicotyl	herb
Acacia	anura	Fabaceae	Angiosperm_Eudicotyl	shrub/tree
Acacia	bivenata	Fabaceae	Angiosperm_Eudicotyl	shrub/tree
Acacia	busilini	Fabaceae	Angiosperm_Eudicotyl	shrub/tree
Acacia	cognata	Fabaceae	Angiosperm_Eudicotyl	shrub/tree
Acacia	dealbata	Fabaceae	Angiosperm_Eudicotyl	shrub/tree
Acacia	elata	Fabaceae	Angiosperm_Eudicotyl	tree
Acacia	implexa	Fabaceae	Angiosperm_Eudicotyl	shrub/tree
Acacia	irrorata	Fabaceae	Angiosperm_Eudicotyl	shrub/tree
Acacia	jonesii	Fabaceae	Angiosperm_Eudicotyl	shrub
Acacia	koa	Fabaceae	Angiosperm_Eudicotyl	tree
Acacia	ligulata	Fabaceae	Angiosperm_Eudicotyl	shrub/tree
Acacia	mearnsii	Fabaceae	Angiosperm_Eudicotyl	tree
Acacia	melanoxylon	Fabaceae	Angiosperm_Eudicotyl	tree
Acacia	penulata	Fabaceae	Angiosperm_Eudicotyl	tree

Results

Number of Nodes: 739

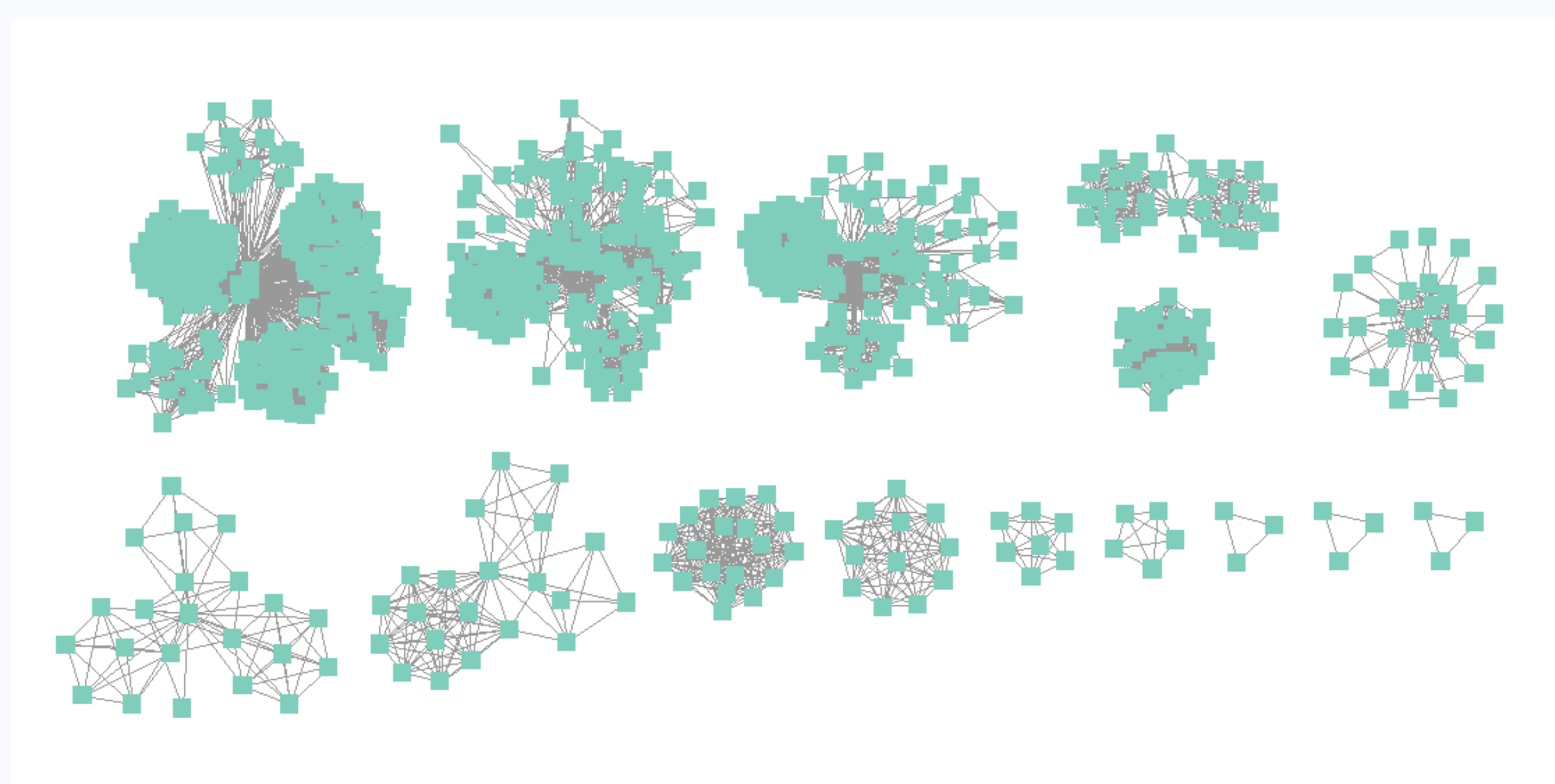
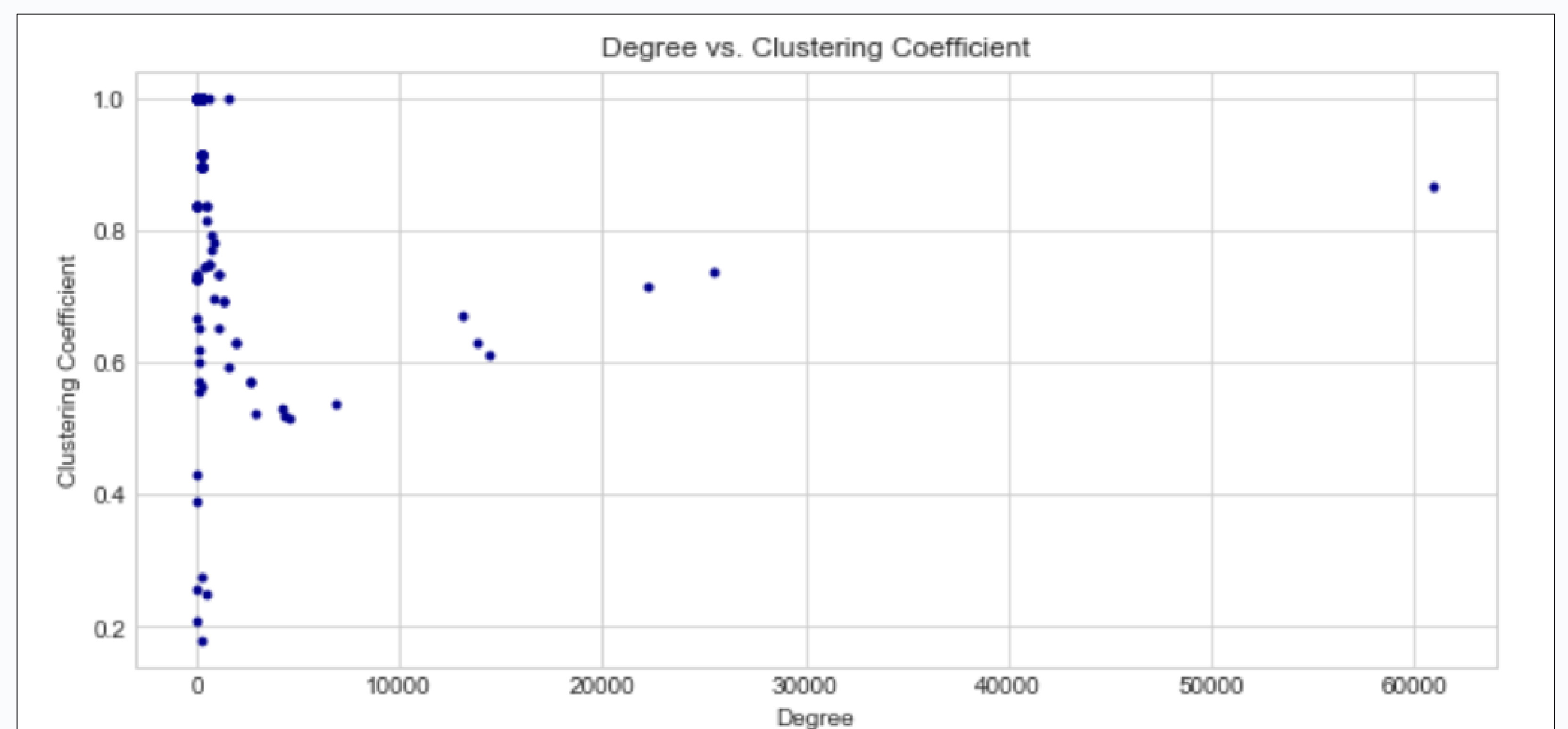
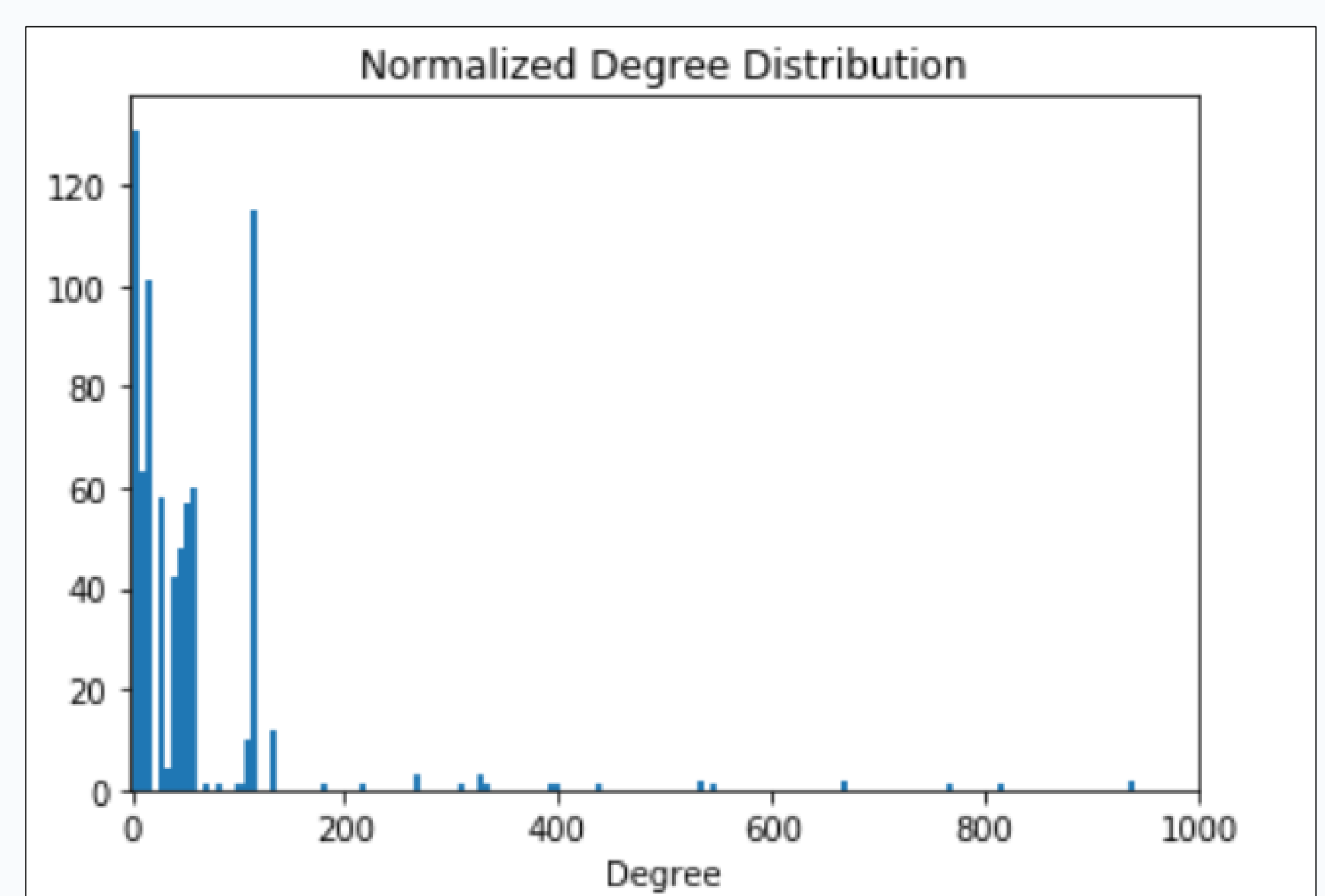
Number of Links: 131,321

Connected Components: 15

Clustering Coefficient: 0.986

Smallest Degree: 2

Lowest Degree: 60,950



Discussion

For our plant taxonomic network projection, the number of nodes is the sum of the families, genera, and species in the taxonomic group, Gymnosperms, thus is exactly what was expected. The number of possible links for 739 nodes follows $\frac{N(N-1)}{2} = 272,691$, whereas the number of links is 131,321. This value makes is what one might expect, since the nodes from different families (components) are by definition not able to form a link. There are 15 connected components, just what one might predict since there are 15 families in our dataset. The clustering coefficient is very close to 1. This is what one would expect for each component is very near fully connected. Since the network is unweighted, the degree distribution does not give us the most useful information.

Moving Forward

We should transform this unweighted, projected network to a weighted network. To as the original question, does the taxonomic classification correlate with genetic classifications, we need to compare the results of the analysis of the weighted network to a dataset with ground truth data on genetic characteristics.

Acknowledgments

The *TRY* initiative and data base is hosted, developed and maintained by J. Kattge and G. Bönisch (Max Planck Institute for Biogeochemistry, Jena, Germany).
Cytoscape Program

Plant Taxonomic Network Projection

Chloe T. Calderon

Department of Physics and Astronomy - Northwestern University

Abstract

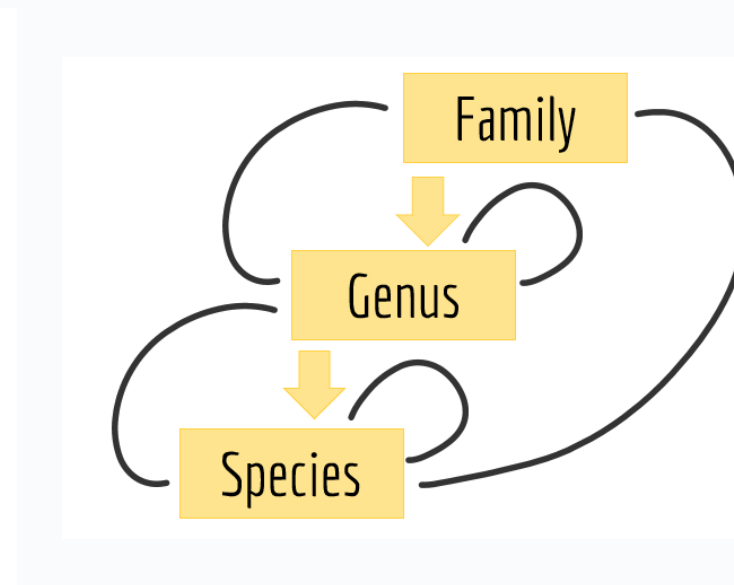
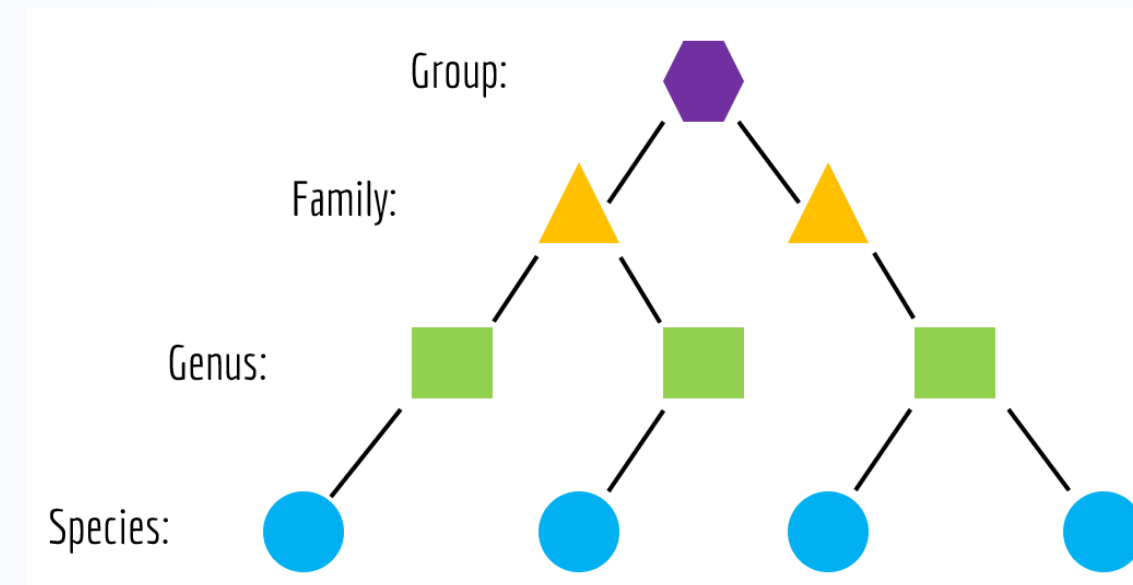
Plant taxonomic classification is based on observable, macroscopic shared characteristics.

What would a classification system based on genetic characteristics (number of genes, number of copies, etc.) look like? How do genetic characteristics relate to our current classification system?

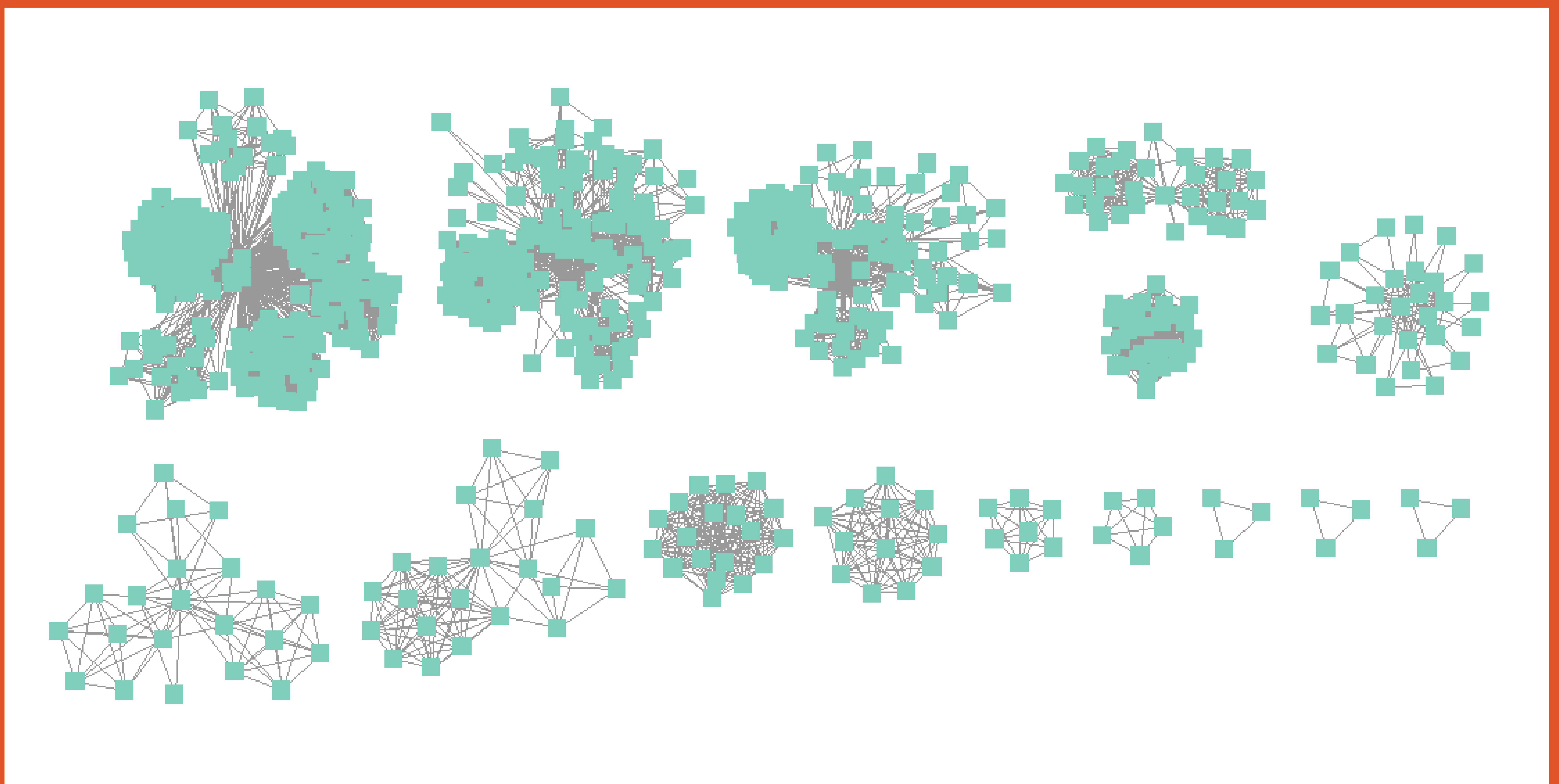
We begin to answer these questions studying the tripartite network for the Gymnosperm taxonomic group as a single network. The original network has a hierarchical structure that is difficult to gain information from, thus we will project this network. The three kinds of nodes for this tripartite network are the taxonomic family, genus, and species.

Method

We began with data from the *TRY Plant Trait Database*. There were 66,043 species from ~10 groups in the database. Our analysis was narrowed to one group, the Gymnosperms, plants with unprotected seeds including conifers, cycads, and ginkos. In this group, there are 15 families, 77 genera, and 647 species. Links were formed between species in the same genus, genera in the same family, species and its genera, genera and its family, and species and its family. Cytoscape was used to analyze this projection network.

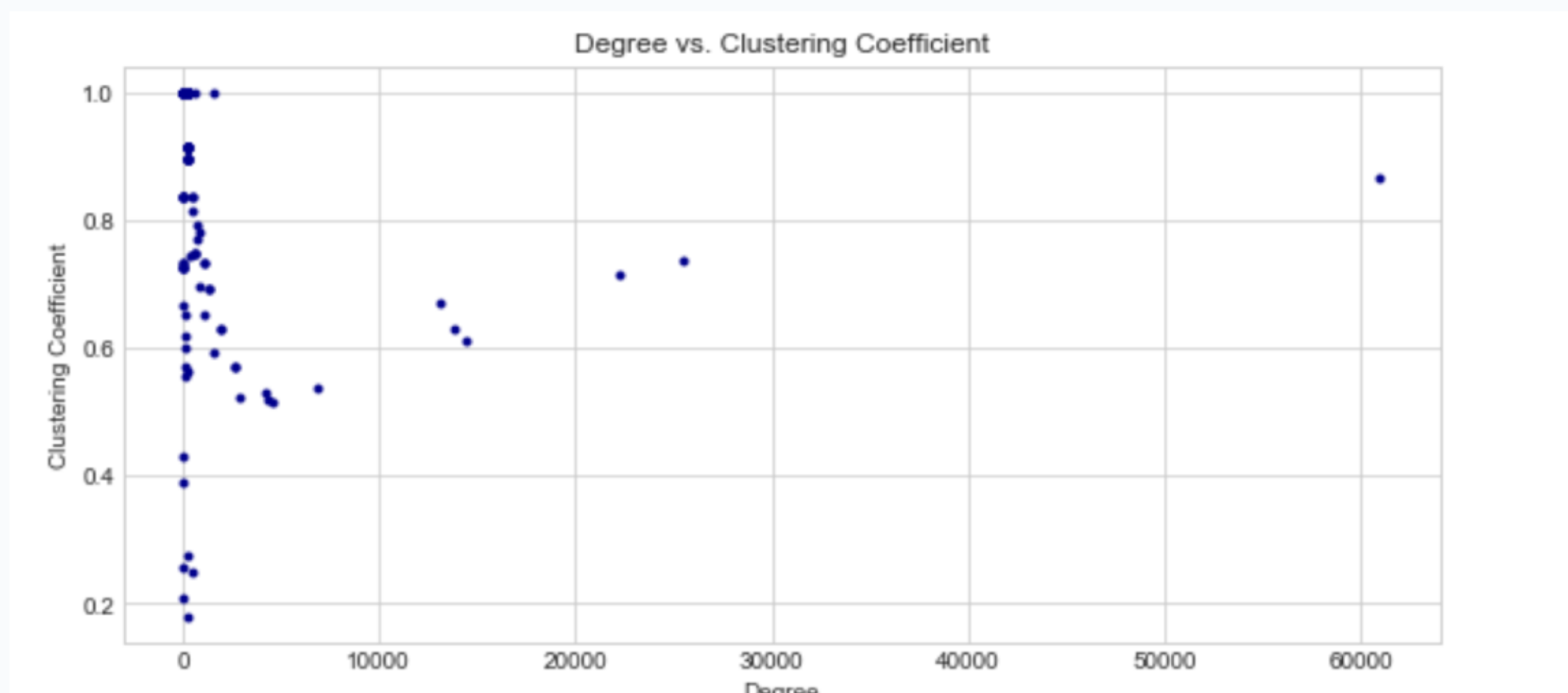
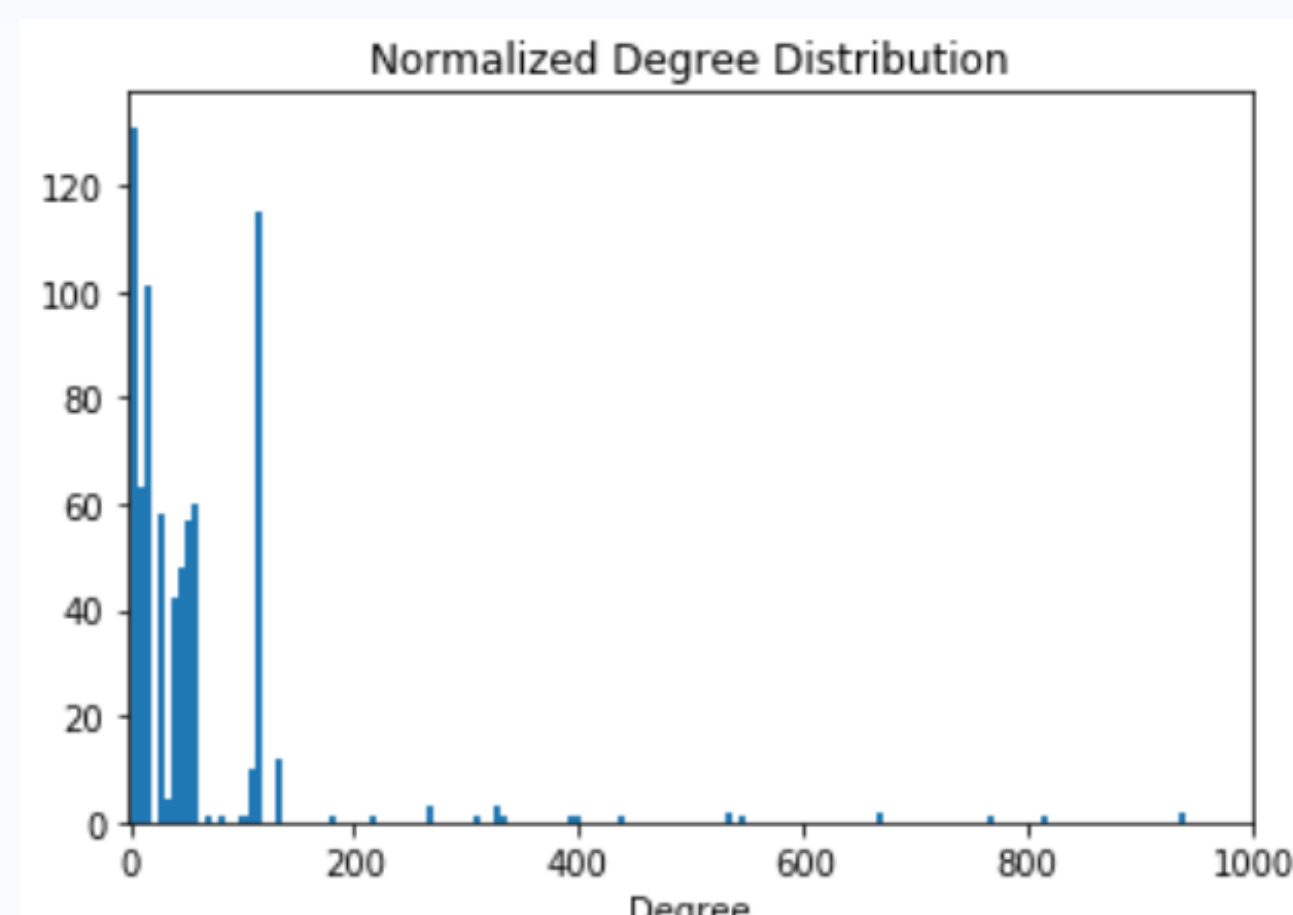


	D	E	F	G	H
1	Genus	Species/epithet	Family	PhylogeneticGroup	PlantGrowthForm
2	Abrotia	maritima	Nyctaginaceae	Angiosperm_Eudicotyl	herb
3	Adiantum	theophrasti	Mahoeaceae	Angiosperm_Eudicotyl	herb
4	Acacia	anura	Fabaceae	Angiosperm_Eudicotyl	shrub/tree
5	Acacia	bivenata	Fabaceae	Angiosperm_Eudicotyl	shrub/tree
6	Acacia	buschii	Fabaceae	Angiosperm_Eudicotyl	shrub/tree
7	Acacia	cognata	Fabaceae	Angiosperm_Eudicotyl	shrub/tree
8	Acacia	dealbata	Fabaceae	Angiosperm_Eudicotyl	shrub/tree
9	Acacia	elata	Fabaceae	Angiosperm_Eudicotyl	tree
10	Acacia	implicata	Fabaceae	Angiosperm_Eudicotyl	shrub/tree
11	Acacia	irrorata	Fabaceae	Angiosperm_Eudicotyl	shrub/tree
12	Acacia	jonesii	Fabaceae	Angiosperm_Eudicotyl	shrub
13	Acacia	koa	Fabaceae	Angiosperm_Eudicotyl	tree
14	Acacia	ligulata	Fabaceae	Angiosperm_Eudicotyl	shrub/tree
15	Acacia	mearnsii	Fabaceae	Angiosperm_Eudicotyl	tree
16	Acacia	melanoxylon	Fabaceae	Angiosperm_Eudicotyl	tree
17	Acacia	penulata	Fabaceae	Angiosperm_Eudicotyl	tree



Results

Number of Nodes: 739
 Number of Links: 131,321
 Connected Components: 15
 Clustering Coefficient: 0.986
 Smallest Degree: 2
 Lowest Degree: 60,950



Discussion

For our plant taxonomic network projection, the number of nodes is the sum of the families, genera, and species in the taxonomic group, Gymnosperms, thus is exactly what was expected. The number of possible links for 739 nodes follows $\frac{N(N-1)}{2} = 272,691$, whereas the number of links is 131,321. This value makes is what one might expect, since the nodes from different families (components) are by definition not able to form a link. There are 15 connected components, just what one might predict since there are 15 families in our dataset. The clustering coefficient is very close to 1. This is what one would expect for each component is very near fully connected. Since the network is unweighted, the degree distribution does not give us the most useful information.

Moving Forward

We should transform this unweighted, projected network to a weighted network. To as the original question, does the taxonomic classification correlate with genetic classifications, we need to compare the results of the analysis of the weighted network to a dataset with ground truth data on genetic characteristics.

Acknowledgments

The *TRY* initiative and data base is hosted, developed and maintained by J. Kattge and G. Bönisch (Max Planck Institute for Biogeochemistry, Jena, Germany).
 Cytoscape Program

Plant Taxonomic Network Projection

Chloe T. Calderon

Department of Physics and Astronomy - Northwestern University

Abstract

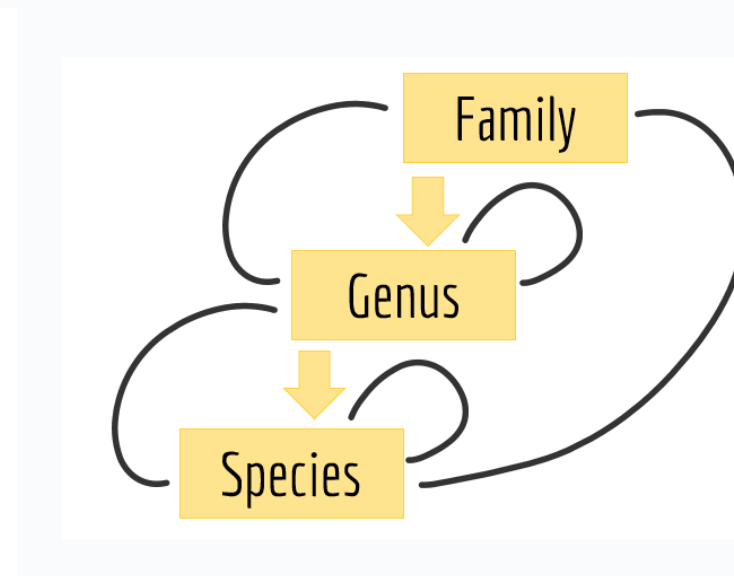
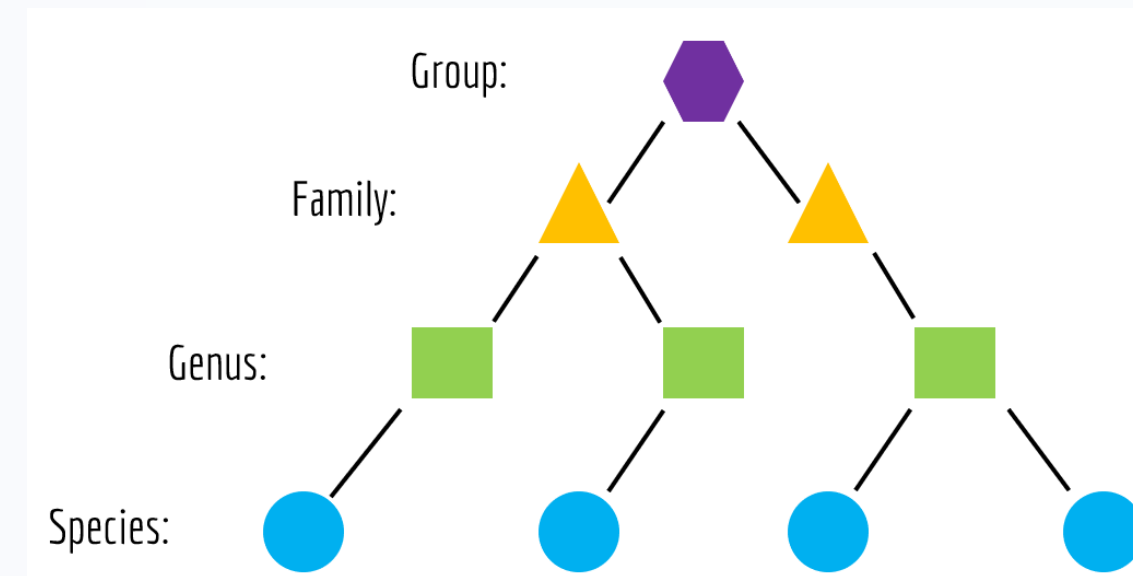
Plant taxonomic classification is based on observable, macroscopic shared characteristics.

What would a classification system based on genetic characteristics (number of genes, number of copies, etc.) look like? How do genetic characteristics relate to our current classification system?

We begin to answer these questions studying the tripartite network for the Gymnosperm taxonomic group as a single network. The original network has a hierarchical structure that is difficult to gain information from, thus we will project this network. The three kinds of nodes for this tripartite network are the taxonomic family, genus, and species.

Method

We began with data from the *TRY Plant Trait Database*. There were 66,043 species from ~10 groups in the database. Our analysis was narrowed to one group, the Gymnosperms, plants with unprotected seeds including conifers, cycads, and ginkos. In this group, there are 15 families, 77 genera, and 647 species. Links were formed between species in the same genus, genera in the same family, and species and its family. Cytoscape was used to analyze this projection network.



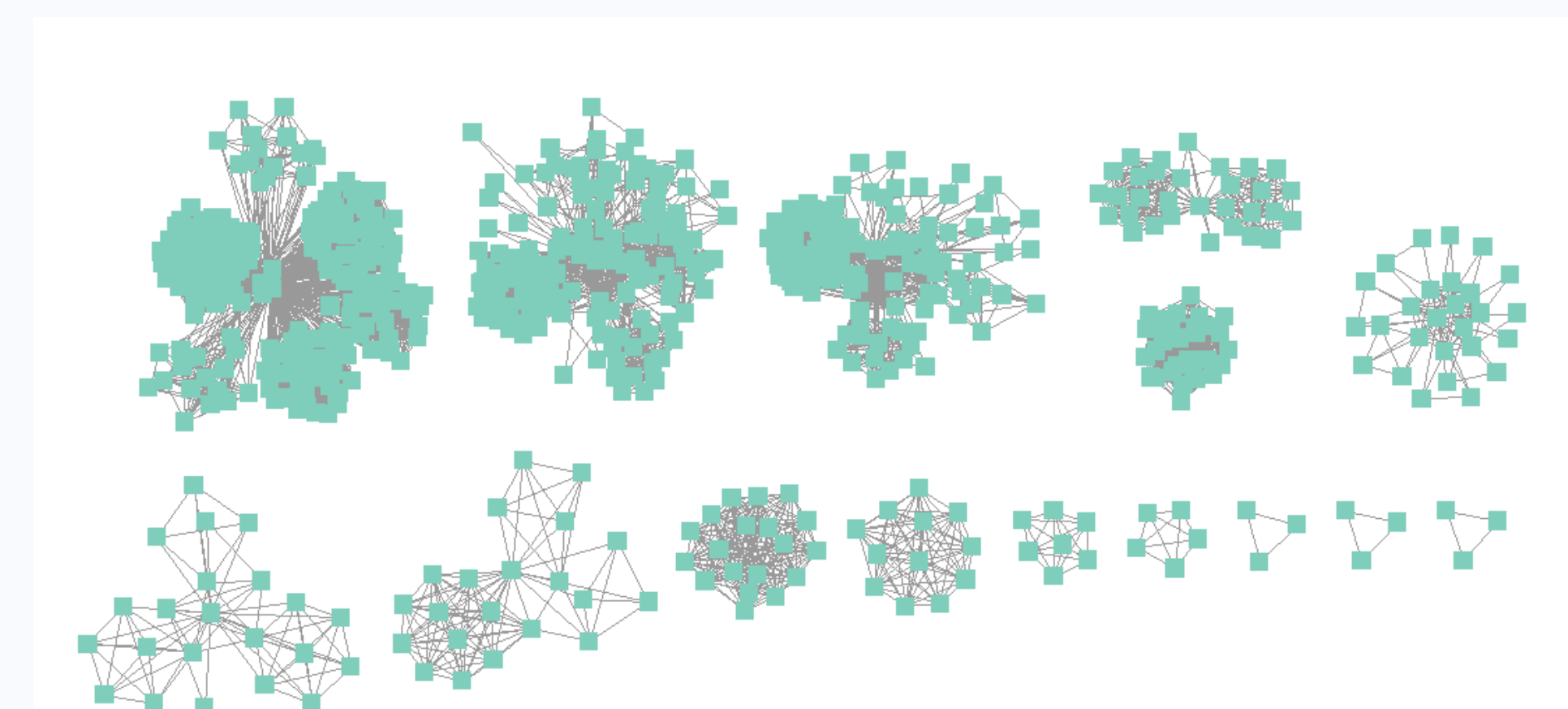
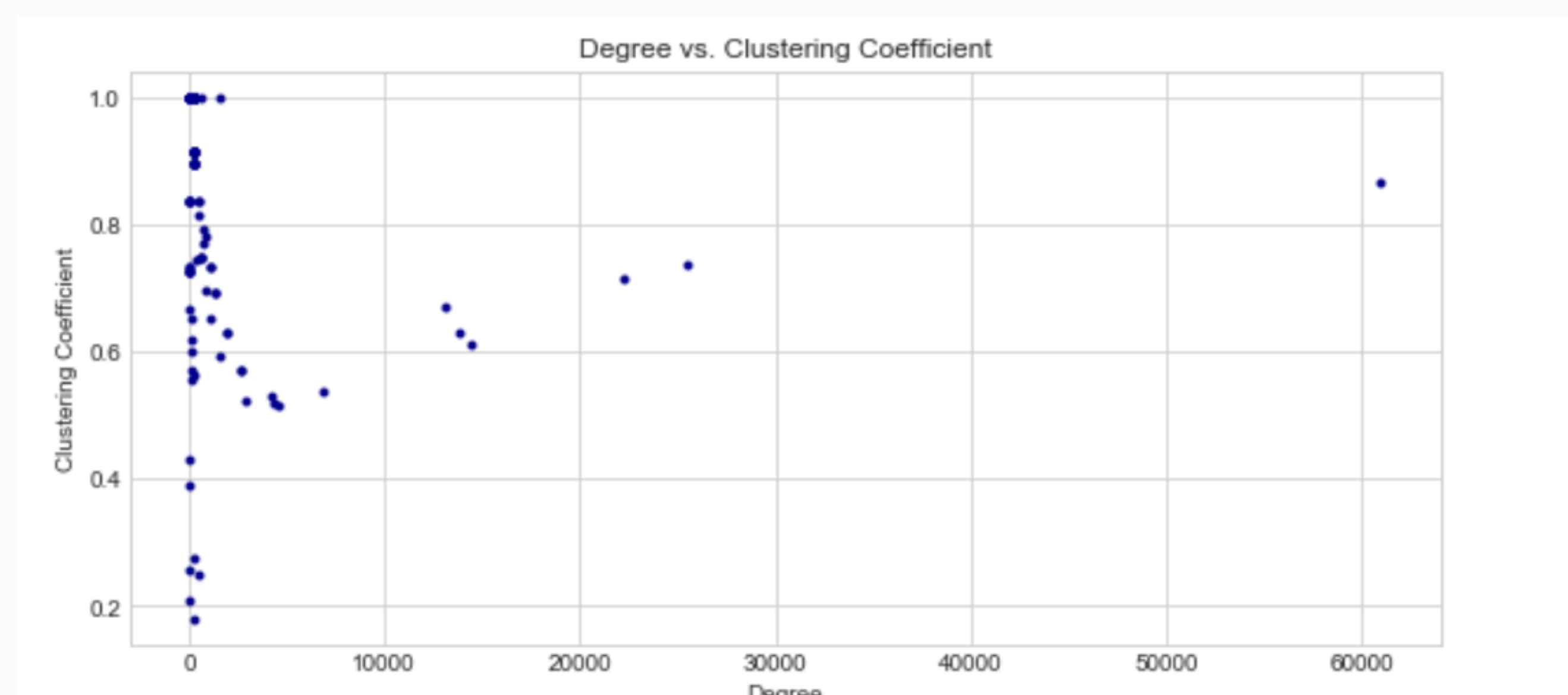
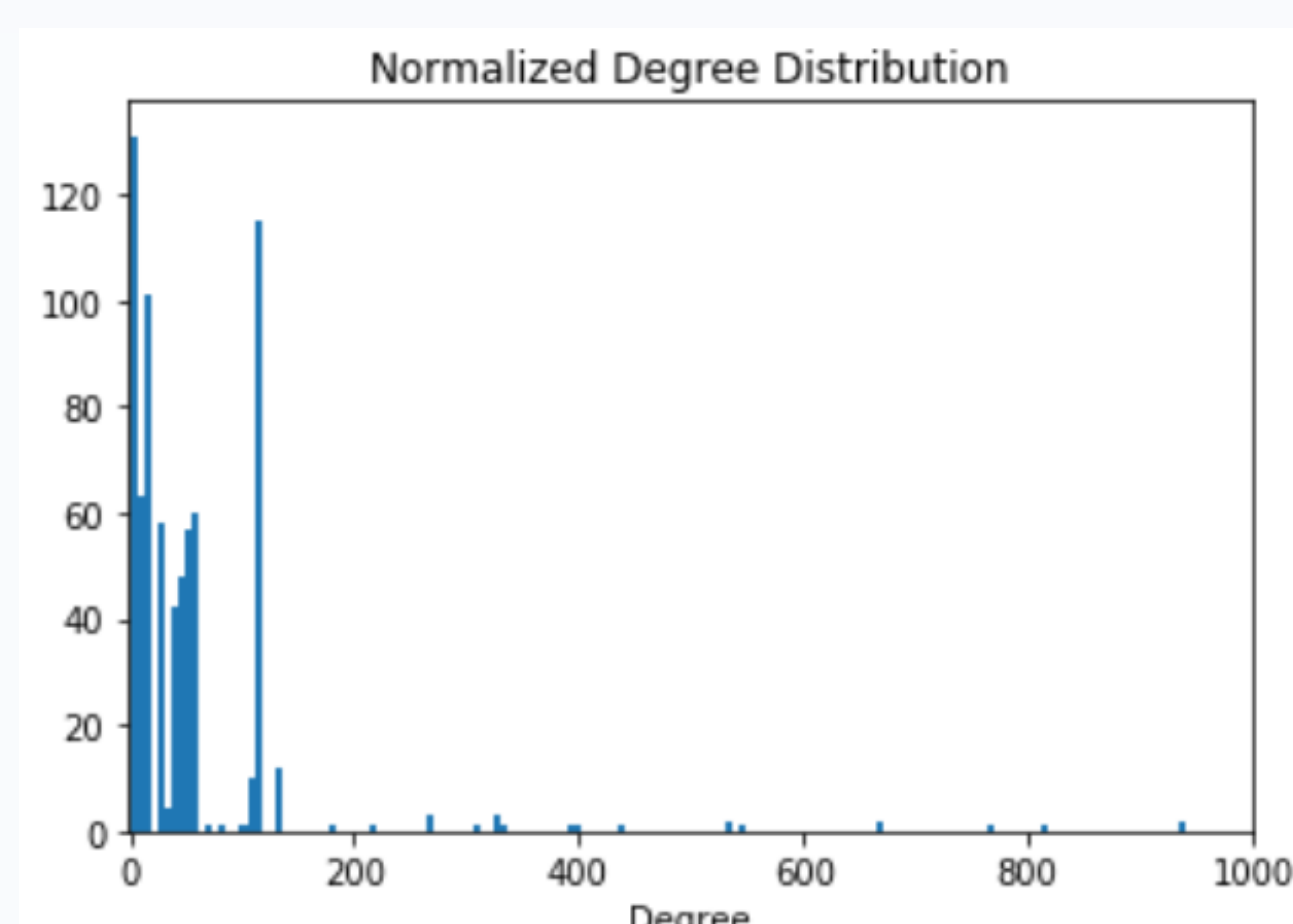
	D	E	F	G	H
1	Genus	Species/Ethet	Family	PhylogeneticGroup	PlantGrowthForm
2	Abronia	maritima	Nyctaginaceae	Angiosperm_Eudicotyl	herb
3	Adiantum	theophrasti	Mahoeaceae	Angiosperm_Eudicotyl	herb
4	Acacia	anura	Fabaceae	Angiosperm_Eudicotyl	shrub/tree
5	Acacia	bivennata	Fabaceae	Angiosperm_Eudicotyl	shrub/tree
6	Acacia	buschii	Fabaceae	Angiosperm_Eudicotyl	shrub/tree
7	Acacia	cognata	Fabaceae	Angiosperm_Eudicotyl	shrub/tree
8	Acacia	deubata	Fabaceae	Angiosperm_Eudicotyl	shrub/tree
9	Acacia	elata	Fabaceae	Angiosperm_Eudicotyl	tree
10	Acacia	implicata	Fabaceae	Angiosperm_Eudicotyl	shrub/tree
11	Acacia	irrorata	Fabaceae	Angiosperm_Eudicotyl	shrub/tree
12	Acacia	jonesii	Fabaceae	Angiosperm_Eudicotyl	shrub
13	Acacia	koa	Fabaceae	Angiosperm_Eudicotyl	tree
14	Acacia	ligulata	Fabaceae	Angiosperm_Eudicotyl	shrub/tree
15	Acacia	mearnsii	Fabaceae	Angiosperm_Eudicotyl	tree
16	Acacia	melanoxylon	Fabaceae	Angiosperm_Eudicotyl	tree
17	Acacia	pendula	Fabaceae	Angiosperm_Eudicotyl	tree

Discussion

For our plant taxonomic network projection, the number of nodes is the sum of the families, genera, and species in the taxonomic group, Gymnosperms, thus is exactly what was expected. The number of possible links for 739 nodes follows $\frac{N(N-1)}{2} = 272,691$, whereas the number of links is 131,321. This value makes is what one might expect, since the nodes from different families (components) are by definition not able to form a link. There are 15 connected components, just what one might predict since there are 15 families in our dataset. The clustering coefficient is very close to 1. This is what one would expect for each component is very near fully connected. Since the network is unweighted, the degree distribution does not give us the most useful information.

Results

Number of Nodes: 739
Number of Links: 131,321
Connected Components: 15
Clustering Coefficient: 0.986
Smallest Degree: 2
Lowest Degree: 60,950



Moving Forward

We should transform this unweighted, projected network to a weighted network. To as the original question, does the taxonomic classification correlate with genetic classifications, we need to compare the results of the analysis of the weighted network to a dataset with ground truth data on genetic characteristics.

Acknowledgments

The *TRY* initiative and data base is hosted, developed and maintained by J. Kattge and G. Bönisch (Max Planck Institute for Biogeochemistry, Jena, Germany).
Cytoscape Program

Plant Taxonomic Network Projection

Chloe T. Calderon

Department of Physics and Astronomy - Northwestern University

Abstract

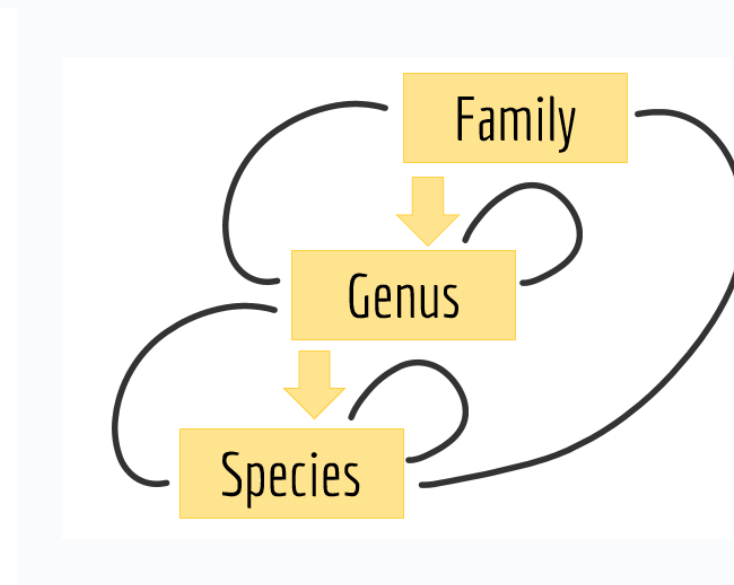
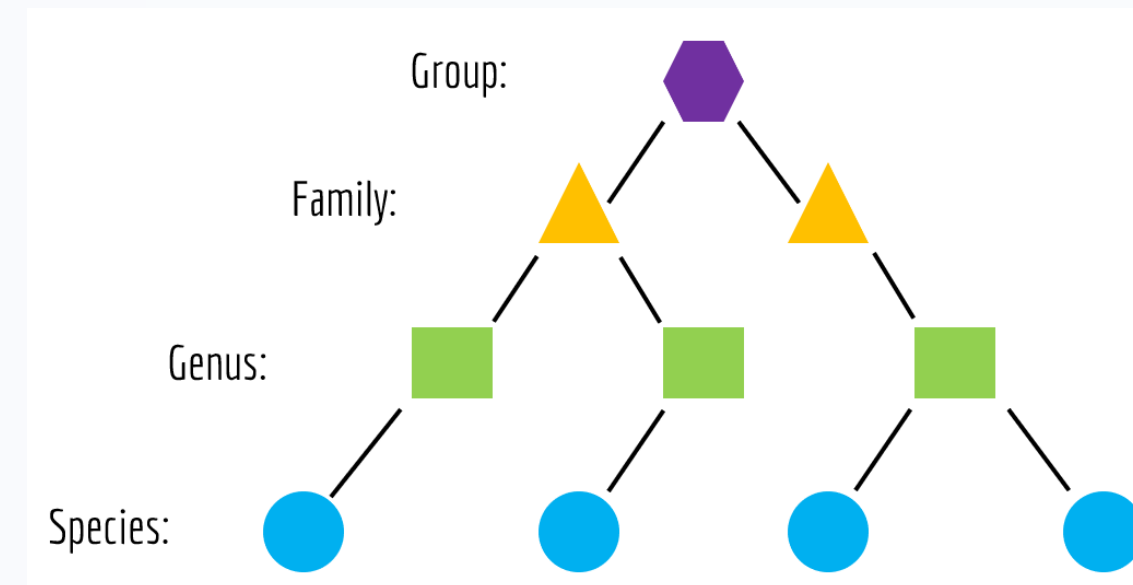
Plant taxonomic classification is based on observable, macroscopic shared characteristics.

What would a classification system based on genetic characteristics (number of genes, number of copies, etc.) look like? How do genetic characteristics relate to our current classification system?

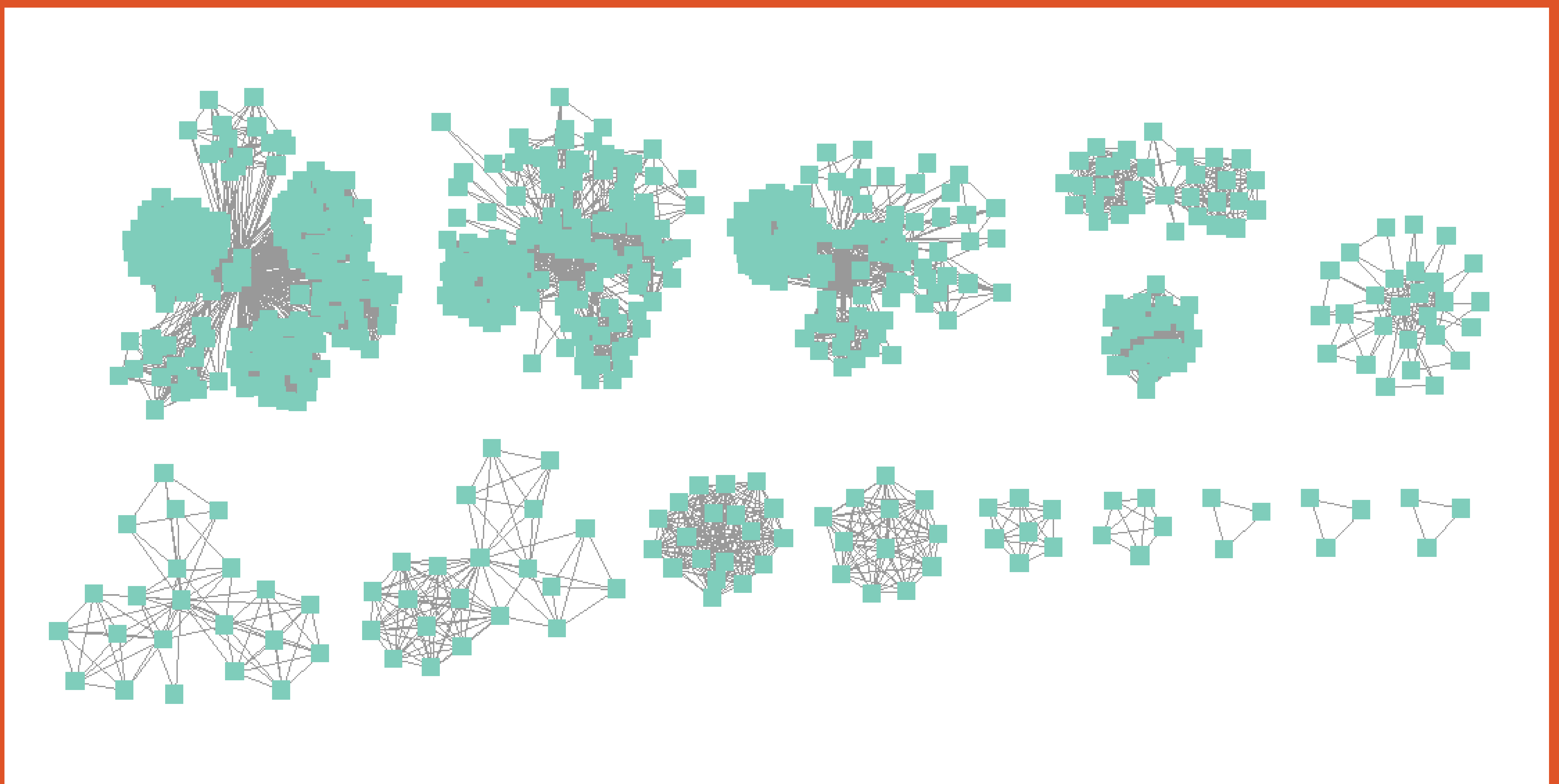
We begin to answer these questions studying the tripartite network for the Gymnosperm taxonomic group as a single network. The original network has a hierarchical structure that is difficult to gain information from, thus we will project this network. The three kinds of nodes for this tripartite network are the taxonomic family, genus, and species.

Method

We began with data from the *TRY Plant Trait Database*. There were 66,043 species from ~10 groups in the database. Our analysis was narrowed to one group, the Gymnosperms, plants with unprotected seeds including conifers, cycads, and ginkos. In this group, there are 15 families, 77 genera, and 647 species. Links were formed between species in the same genus, genera in the same family, species and its genera, genera and its family, and species and its family. Cytoscape was used to analyze this projection network.

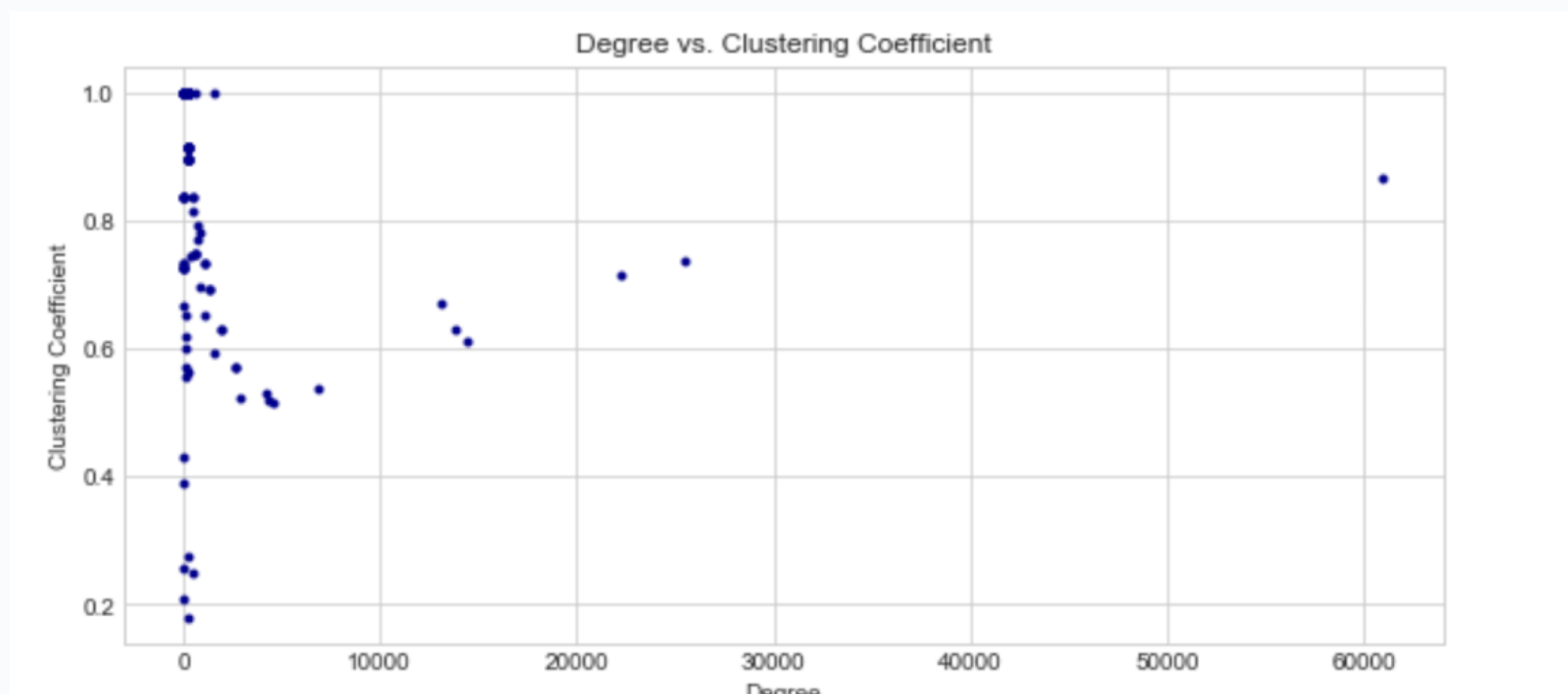
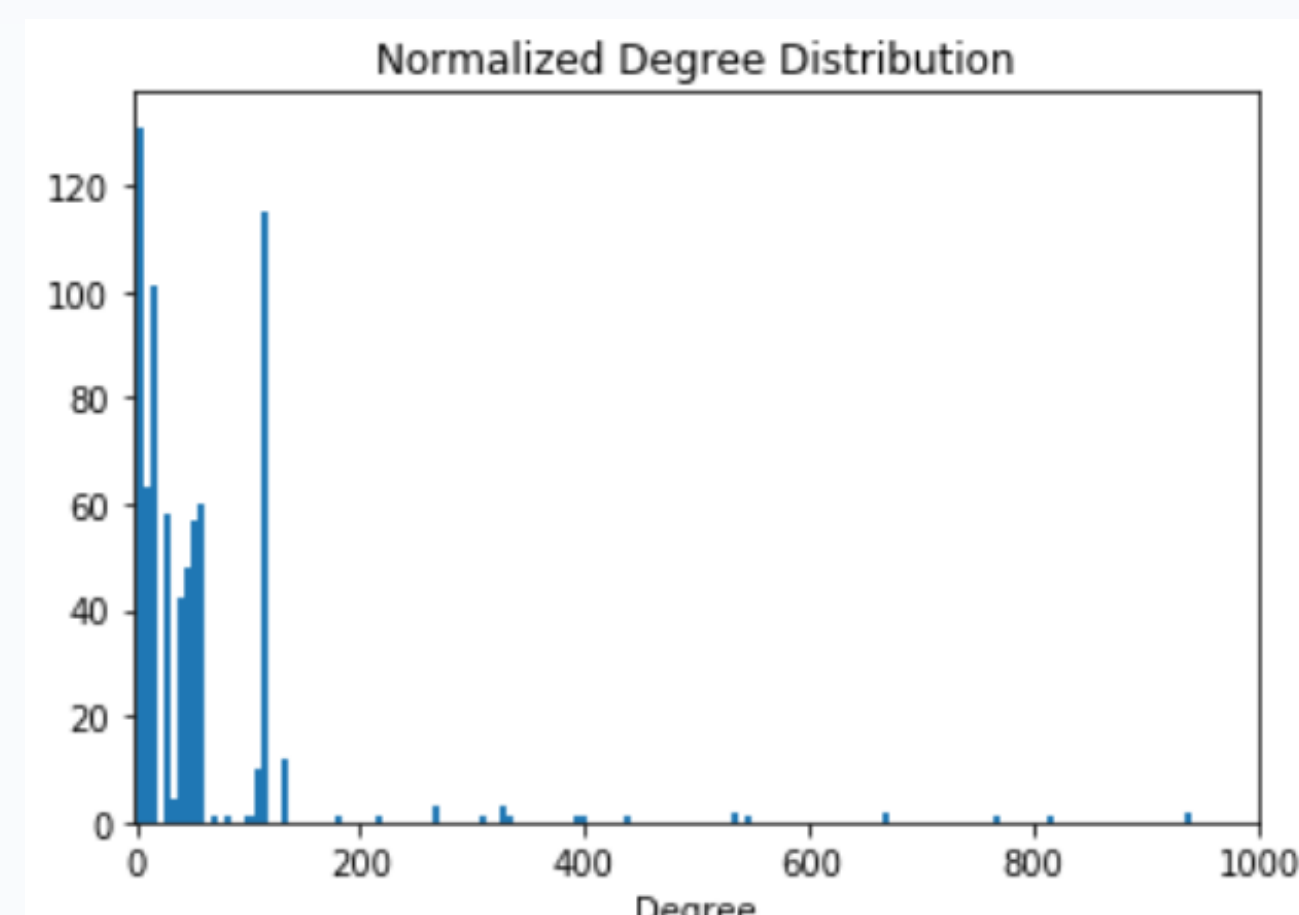


	D	E	F	G	H
1	Genus	Species/epithet	Family	PhylogeneticGroup	PlantGrowthForm
2	Abrotia	maritima	Nyctaginaceae	Angiosperm_Eudicotyl	herb
3	Abutilon	theophrasti	Malvaceae	Angiosperm_Eudicotyl	herb
4	Acacia	anura	Fabaceae	Angiosperm_Eudicotyl	shrub/tree
5	Acacia	bivennata	Fabaceae	Angiosperm_Eudicotyl	shrub/tree
6	Acacia	buschii	Fabaceae	Angiosperm_Eudicotyl	shrub/tree
7	Acacia	cognata	Fabaceae	Angiosperm_Eudicotyl	shrub/tree
8	Acacia	deubata	Fabaceae	Angiosperm_Eudicotyl	shrub/tree
9	Acacia	elata	Fabaceae	Angiosperm_Eudicotyl	tree
10	Acacia	impinea	Fabaceae	Angiosperm_Eudicotyl	shrub/tree
11	Acacia	irrorata	Fabaceae	Angiosperm_Eudicotyl	shrub/tree
12	Acacia	jonesii	Fabaceae	Angiosperm_Eudicotyl	shrub
13	Acacia	koa	Fabaceae	Angiosperm_Eudicotyl	tree
14	Acacia	ligulata	Fabaceae	Angiosperm_Eudicotyl	shrub/tree
15	Acacia	mearnsii	Fabaceae	Angiosperm_Eudicotyl	tree
16	Acacia	melanoxylon	Fabaceae	Angiosperm_Eudicotyl	tree
17	Acacia	penulata	Fabaceae	Angiosperm_Eudicotyl	tree



Results

Number of Nodes: 739
 Number of Links: 131,321
 Connected Components: 15
 Clustering Coefficient: 0.986
 Smallest Degree: 2
 Lowest Degree: 60,950



Discussion

For our plant taxonomic network projection, the number of nodes is the sum of the families, genera, and species in the taxonomic group, Gymnosperms, thus is exactly what was expected. The number of possible links for 739 nodes follows $\frac{N(N-1)}{2} = 272,691$, whereas the number of links is 131,321. This value makes is what one might expect, since the nodes from different families (components) are by definition not able to form a link. There are 15 connected components, just what one might predict since there are 15 families in our dataset. The clustering coefficient is very close to 1. This is what one would expect for each component is very near fully connected. Since the network is unweighted, the degree distribution does not give us the most useful information.

Moving Forward

We should transform this unweighted, projected network to a weighted network. To as the original question, does the taxonomic classification correlate with genetic classifications, we need to compare the results of the analysis of the weighted network to a dataset with ground truth data on genetic characteristics.

Acknowledgments

The *TRY* initiative and data base is hosted, developed and maintained by J. Kattge and G. Bönisch (Max Planck Institute for Biogeochemistry, Jena, Germany).
 Cytoscape Program

Plant Taxonomic Network Projection

Chloe T. Calderon

Department of Physics and Astronomy - Northwestern University

Abstract

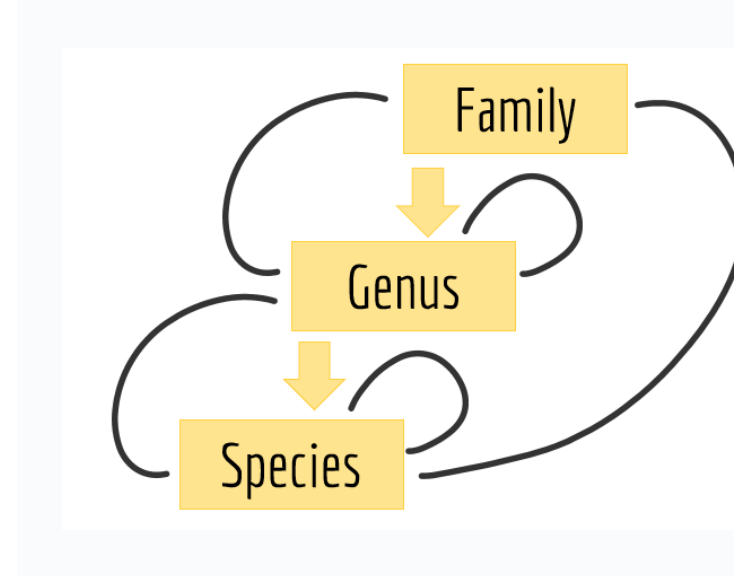
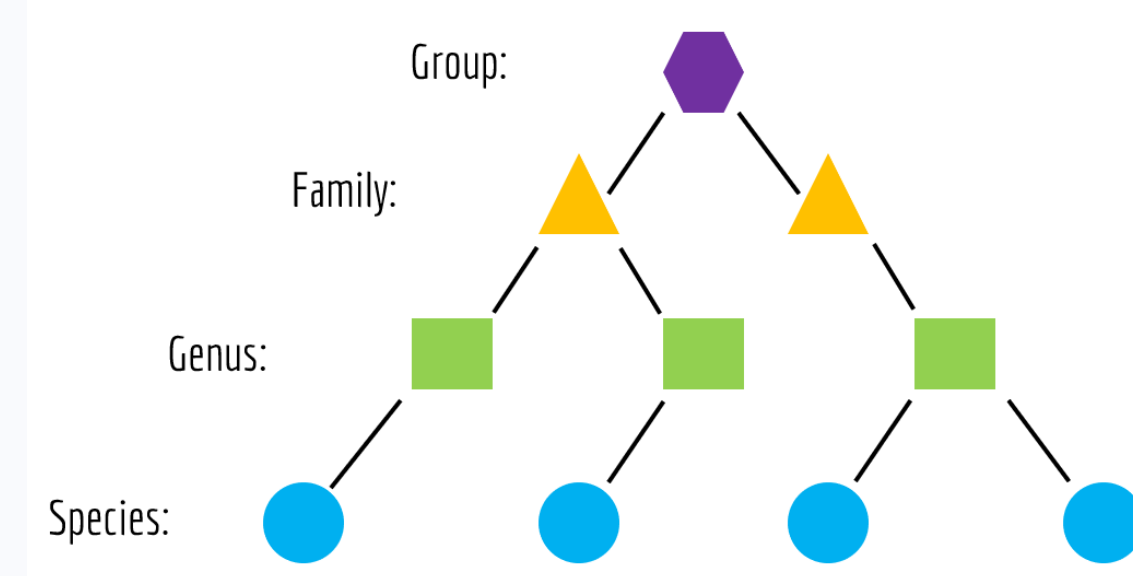
Plant taxonomic classification is based on observable, macroscopic shared characteristics.

What would a classification system based on genetic characteristics (number of genes, number of copies, etc.) look like? How do genetic characteristics relate to our current classification system?

We begin to answer these questions studying the tripartite network for the Gymnosperm taxonomic group onto a single network. The original network has a hierarchical structure that is difficult to gain information from, thus we will project this network. The three kinds of nodes for this tripartite network are the taxonomic family, genus, and species.

Method

We began with data from the *TRY Plant Trait Database*. There were 66,043 species from ~10 groups in the database. Our analysis was narrowed to one group, the Gymnosperms, plants with unprotected seeds including conifers, cycads, and ginkos. In this group, there are 15 families, 77 genera, and 647 species. Links were formed between species in the same genus, genera in the same family, species and its genera, genera and its family, and species and its family. Cytoscape was used to analyze this projection network.



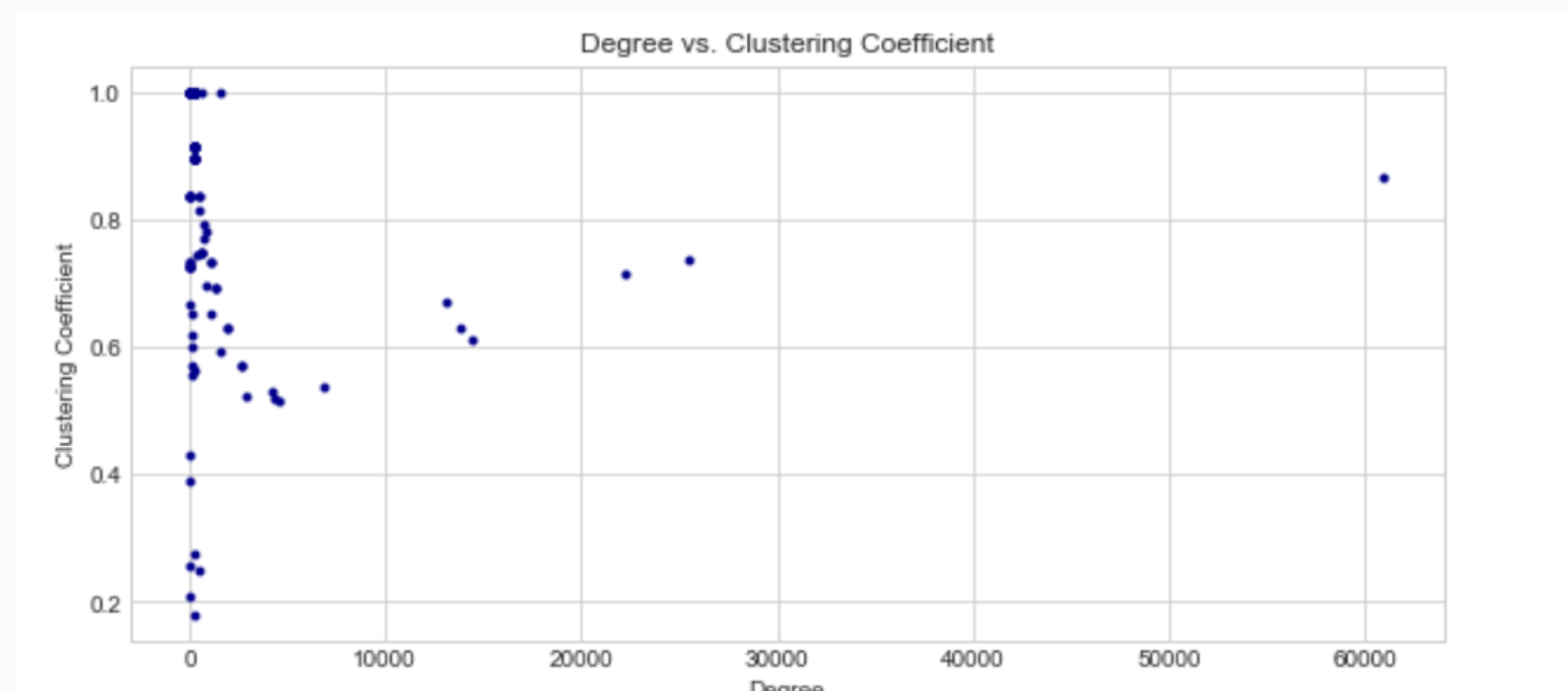
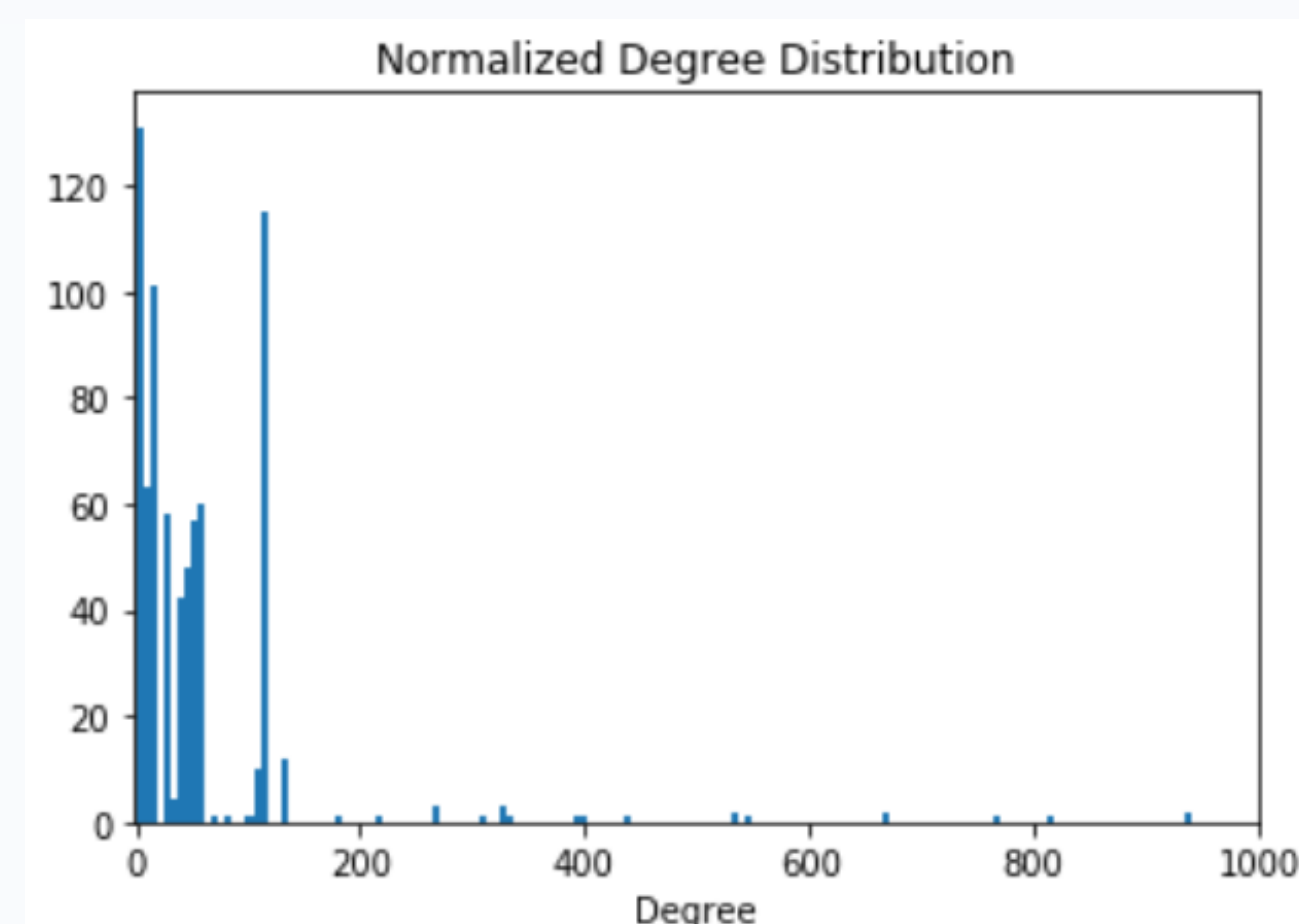
	D	E	F	G	H
1	Genus	Species/Ethet	Family	PhylogeneticGroup	PlantGrowthForm
2	Abronia	maritima	Nyctaginaceae	Angiosperm_Eudicotyl	herb
3	Adiantum	theophrasti	Mahoeaceae	Angiosperm_Eudicotyl	herb
4	Acacia	anura	Fabaceae	Angiosperm_Eudicotyl	shrub/tree
5	Acacia	bivervata	Fabaceae	Angiosperm_Eudicotyl	shrub/tree
6	Acacia	buschii	Fabaceae	Angiosperm_Eudicotyl	shrub/tree
7	Acacia	cognata	Fabaceae	Angiosperm_Eudicotyl	shrub/tree
8	Acacia	dealbata	Fabaceae	Angiosperm_Eudicotyl	shrub/tree
9	Acacia	elata	Fabaceae	Angiosperm_Eudicotyl	tree
10	Acacia	implexa	Fabaceae	Angiosperm_Eudicotyl	shrub/tree
11	Acacia	irrorata	Fabaceae	Angiosperm_Eudicotyl	shrub/tree
12	Acacia	jonesii	Fabaceae	Angiosperm_Eudicotyl	shrub
13	Acacia	koa	Fabaceae	Angiosperm_Eudicotyl	tree
14	Acacia	ligulata	Fabaceae	Angiosperm_Eudicotyl	shrub/tree
15	Acacia	mearnsii	Fabaceae	Angiosperm_Eudicotyl	tree
16	Acacia	melanoxylon	Fabaceae	Angiosperm_Eudicotyl	tree
17	Acacia	pendula	Fabaceae	Angiosperm_Eudicotyl	tree

Moving Forward

We should transform this unweighted, projected network to a weighted network. To answer the original question, does the taxonomic classification correlate with genetic classifications, we need to compare the results of the analysis of the weighted network to a dataset with ground truth data on genetic characteristics.

Results

Number of Nodes: 739
Number of Links: 131,321
Connected Components: 15
Clustering Coefficient: 0.986
Smallest Degree: 2
Lowest Degree: 60,950



Discussion

For our plant taxonomic network projection, the number of nodes is the sum of the families, genera, and species in the taxonomic group, Gymnosperms, thus is exactly what was expected. The number of possible links for 739 nodes follows $\frac{N(N-1)}{2} = 272,691$, whereas the number of links is 131,321. This value makes sense, since the nodes from different families (components) are by definition not able to form a link. There are 15 connected components, just what one might predict since there are 15 families in our dataset. The clustering coefficient is very close to 1. This is what one would expect for each component is very near fully connected. Since the network is unweighted, the degree distribution does not give us the most useful information.

Acknowledgments

The *TRY* initiative and data base is hosted, developed and maintained by J. Kattge and G. Bönisch (Max Planck Institute for Biogeochemistry, Jena, Germany).
Cytoscape Program

Plant Taxonomic Network Projection

Chloe T. Calderon

Department of Physics and Astronomy - Northwestern University

Abstract

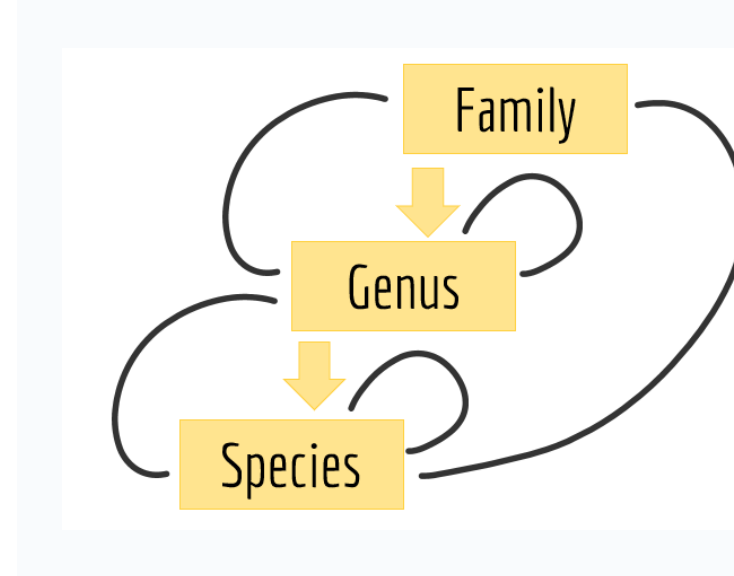
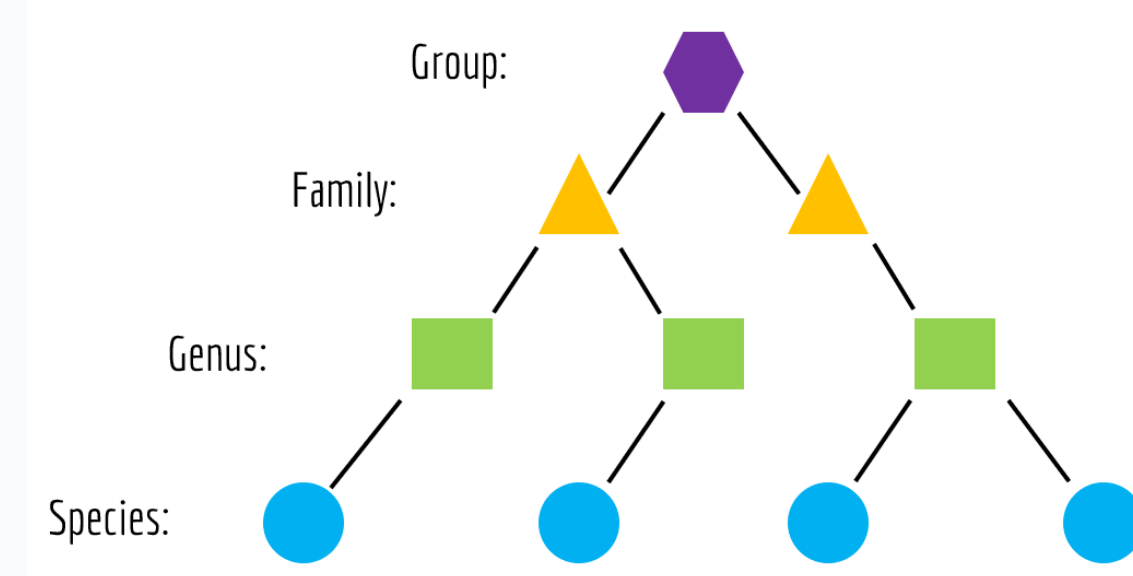
Plant taxonomic classification is based on observable, macroscopic shared characteristics.

What would a classification system based on genetic characteristics (number of genes, number of copies, etc.) look like? How do genetic characteristics relate to our current classification system?

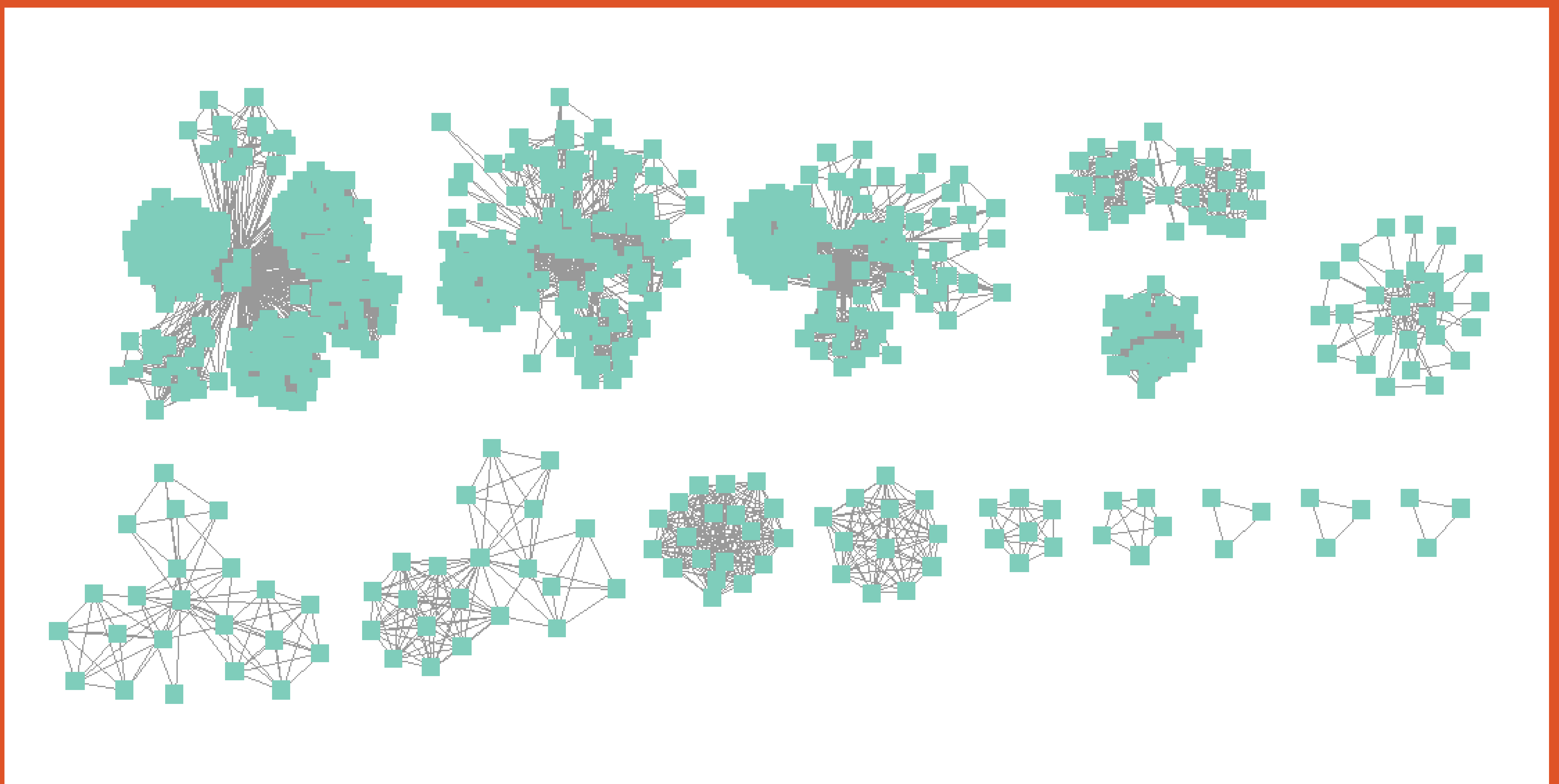
We begin to answer these questions studying the tripartite network for the Gymnosperm taxonomic group as a single network. The original network has a hierarchical structure that is difficult to gain information from, thus we will project this network. The three kinds of nodes for this tripartite network are the taxonomic family, genus, and species.

Method

We began with data from the *TRY Plant Trait Database*. There were 66,043 species from ~10 groups in the database. Our analysis was narrowed to one group, the Gymnosperms, plants with unprotected seeds including conifers, cycads, and ginkos. In this group, there are 15 families, 77 genera, and 647 species. Links were formed between species in the same genus, genera in the same family, species and its genera, genera and its family, and species and its family. Cytoscape was used to analyze this projection network.

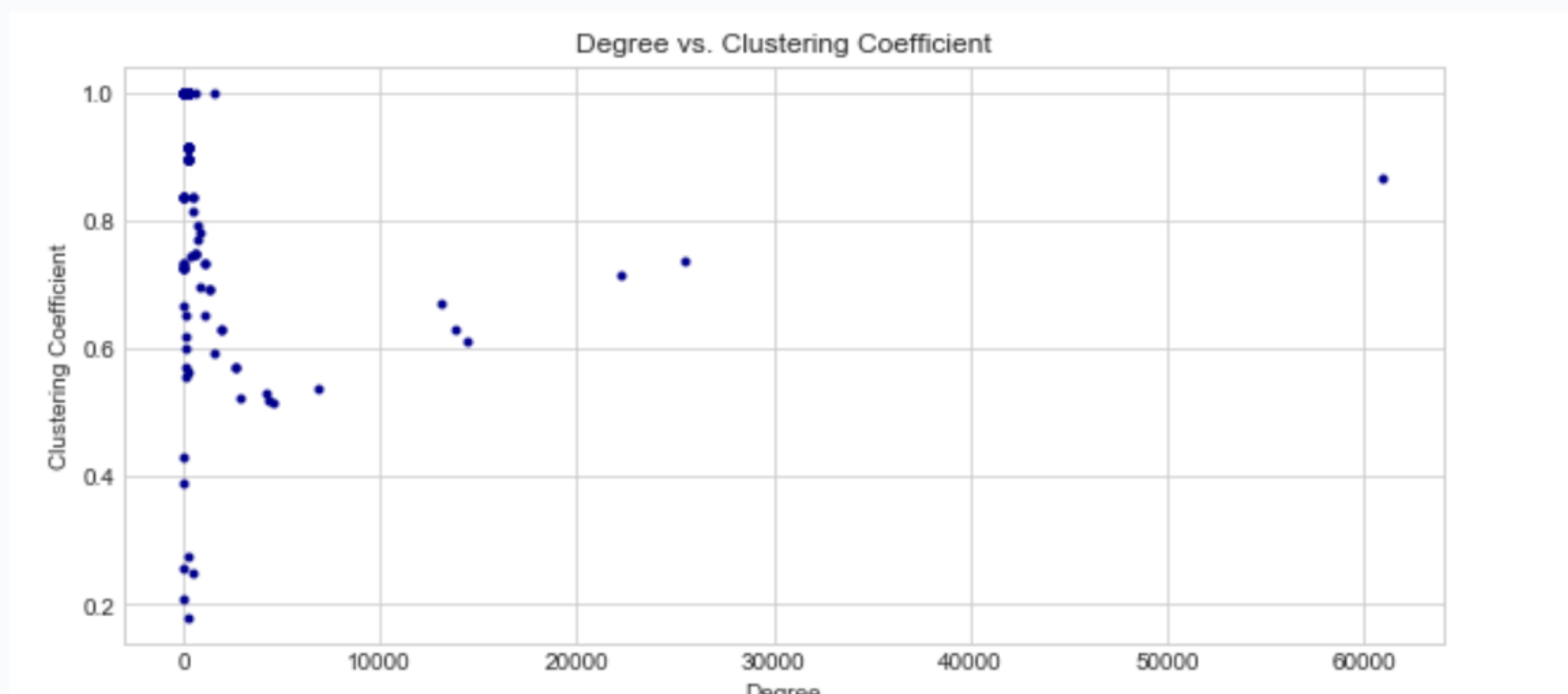
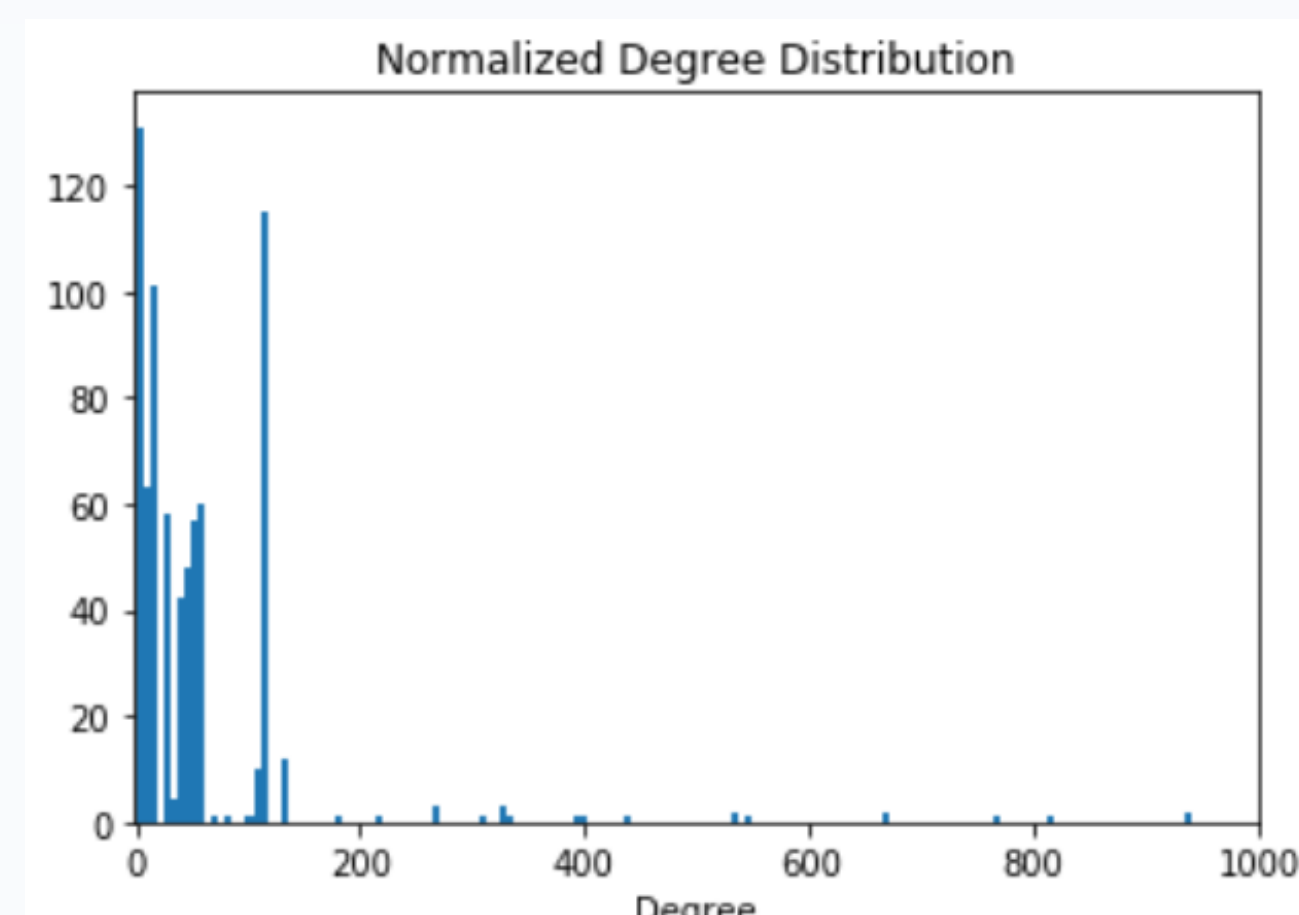


	D	E	F	G	H
1	Genus	Species/Ethet	Family	PhylogeneticGroup	PlantGrowthForm
2	Abrotia	maritima	Nyctaginaceae	Angiosperm_Eudicotyl	herb
3	Abutilon	theophrasti	Malvaceae	Angiosperm_Eudicotyl	herb
4	Acacia	anura	Fabaceae	Angiosperm_Eudicotyl	shrub/tree
5	Acacia	bivenata	Fabaceae	Angiosperm_Eudicotyl	shrub/tree
6	Acacia	busilini	Fabaceae	Angiosperm_Eudicotyl	shrub/tree
7	Acacia	cognata	Fabaceae	Angiosperm_Eudicotyl	shrub/tree
8	Acacia	deubata	Fabaceae	Angiosperm_Eudicotyl	shrub/tree
9	Acacia	elata	Fabaceae	Angiosperm_Eudicotyl	tree
10	Acacia	impinea	Fabaceae	Angiosperm_Eudicotyl	shrub/tree
11	Acacia	irrorata	Fabaceae	Angiosperm_Eudicotyl	shrub/tree
12	Acacia	jonesii	Fabaceae	Angiosperm_Eudicotyl	shrub
13	Acacia	koa	Fabaceae	Angiosperm_Eudicotyl	tree
14	Acacia	ligulata	Fabaceae	Angiosperm_Eudicotyl	shrub/tree
15	Acacia	mearnsii	Fabaceae	Angiosperm_Eudicotyl	tree
16	Acacia	melanoxylon	Fabaceae	Angiosperm_Eudicotyl	tree
17	Acacia	penulata	Fabaceae	Angiosperm_Eudicotyl	tree



Results

Number of Nodes: 739
 Number of Links: 131,321
 Connected Components: 15
 Clustering Coefficient: 0.986
 Smallest Degree: 2
 Lowest Degree: 60,950



Discussion

For our plant taxonomic network projection, the number of nodes is the sum of the families, genera, and species in the taxonomic group, Gymnosperms, thus is exactly what was expected. The number of possible links for 739 nodes follows $\frac{N(N-1)}{2} = 272,691$, whereas the number of links is 131,321. This value makes is what one might expect, since the nodes from different families (components) are by definition not able to form a link. There are 15 connected components, just what one might predict since there are 15 families in our dataset. The clustering coefficient is very close to 1. This is what one would expect for each component is very near fully connected. Since the network is unweighted, the degree distribution does not give us the most useful information.

Moving Forward

We should transform this unweighted, projected network to a weighted network. To as the original question, does the taxonomic classification correlate with genetic classifications, we need to compare the results of the analysis of the weighted network to a dataset with ground truth data on genetic characteristics.

Acknowledgments

The *TRY* initiative and data base is hosted, developed and maintained by J. Kattge and G. Bönisch (Max Planck Institute for Biogeochemistry, Jena, Germany).
 Cytoscape Program