Microbial Network Analysis:

Comparing Node Degree Based on Source Feature

Samuel L. Peoples

Background

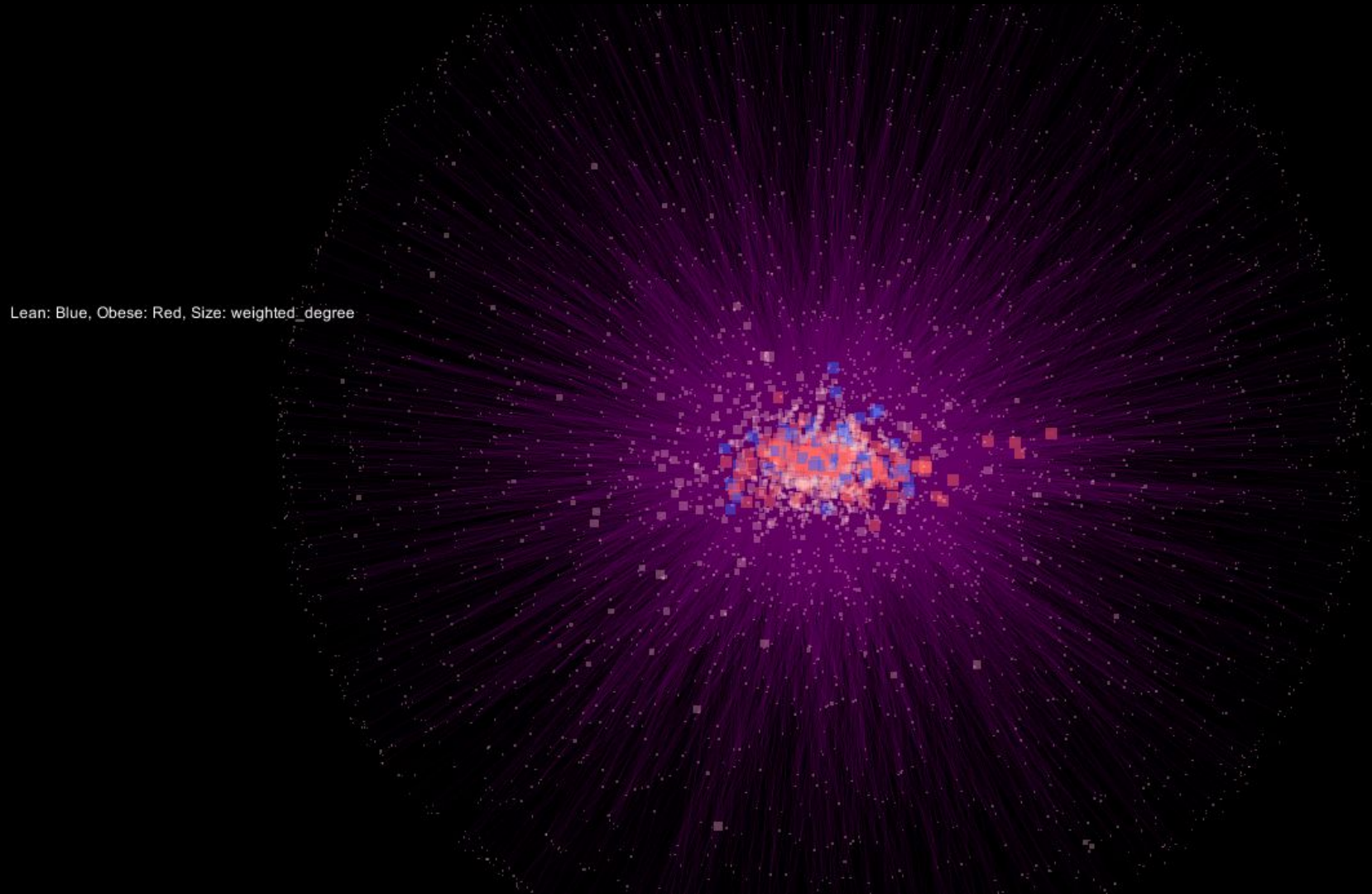
Jeff Gordon's A Core Gut Microbiome in Obese and Lean Twins, drew a handful of conclusions based on a source feature, *obesitycat*, comparing the microbial communities of *Obese*, and *Lean* individuals. These include:

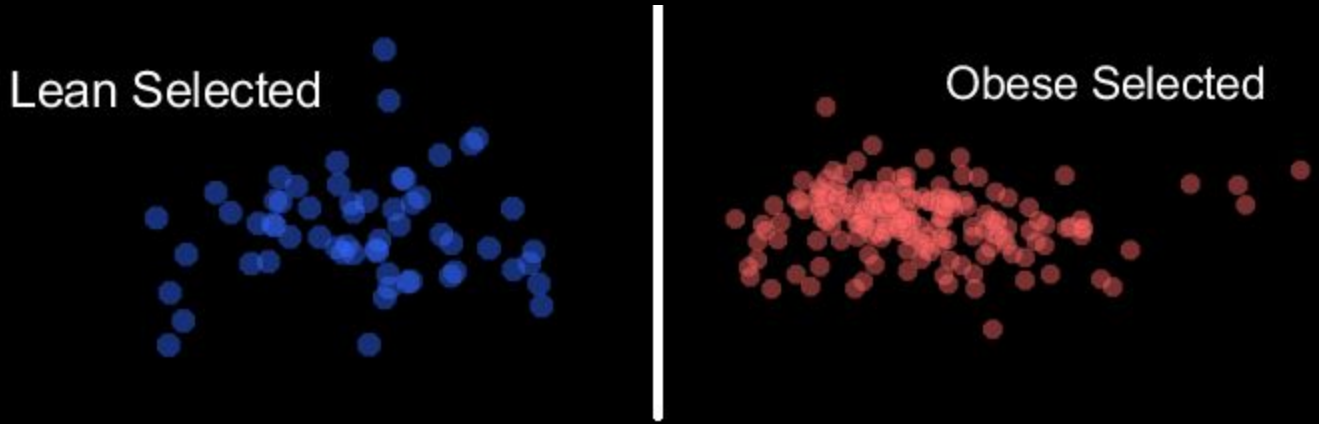
- Wide array of shared genes; there exists a core microbiome at the gene level.
- Obesity is associated with phylum-level changes in the microbiota.
- Deviations from this core microbiome are associated with physiological states.

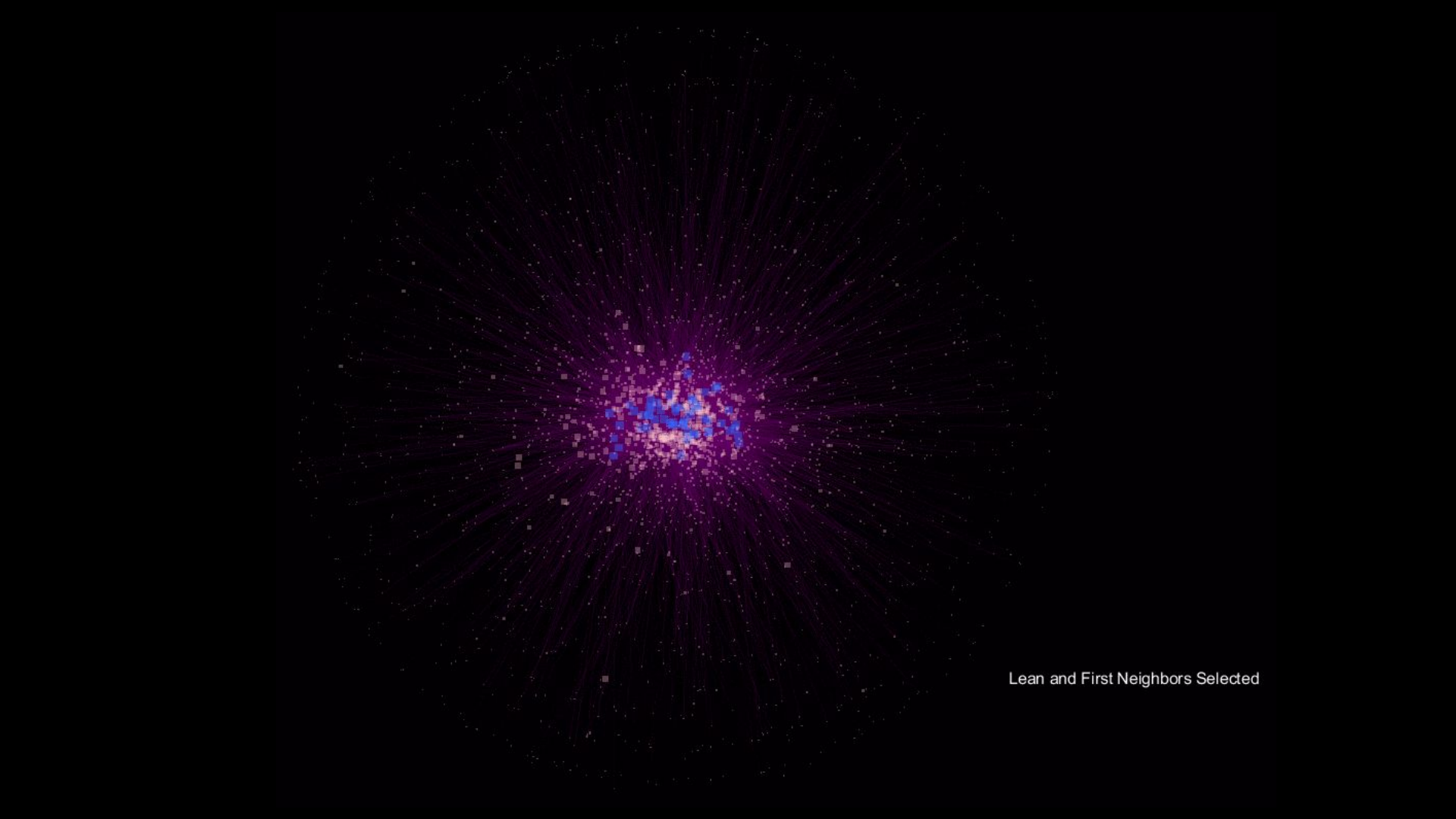
The Network

Using an input *.biom* and mapping file, *make_otu_network.py* was used to generate the Cytoscape network files, which were configured as such:

- Edge-Weighted Spring-Embedded Layout
- Blue: Lean, Red: Obese, Pink: OTUs
- Purple lines: Edges
- Node Size: Weighted Degree


Lean: Blue, Obese: Red, Size: weighted_degree


Lean Selected
Obese Selected


Lean and First Neighbors Selected

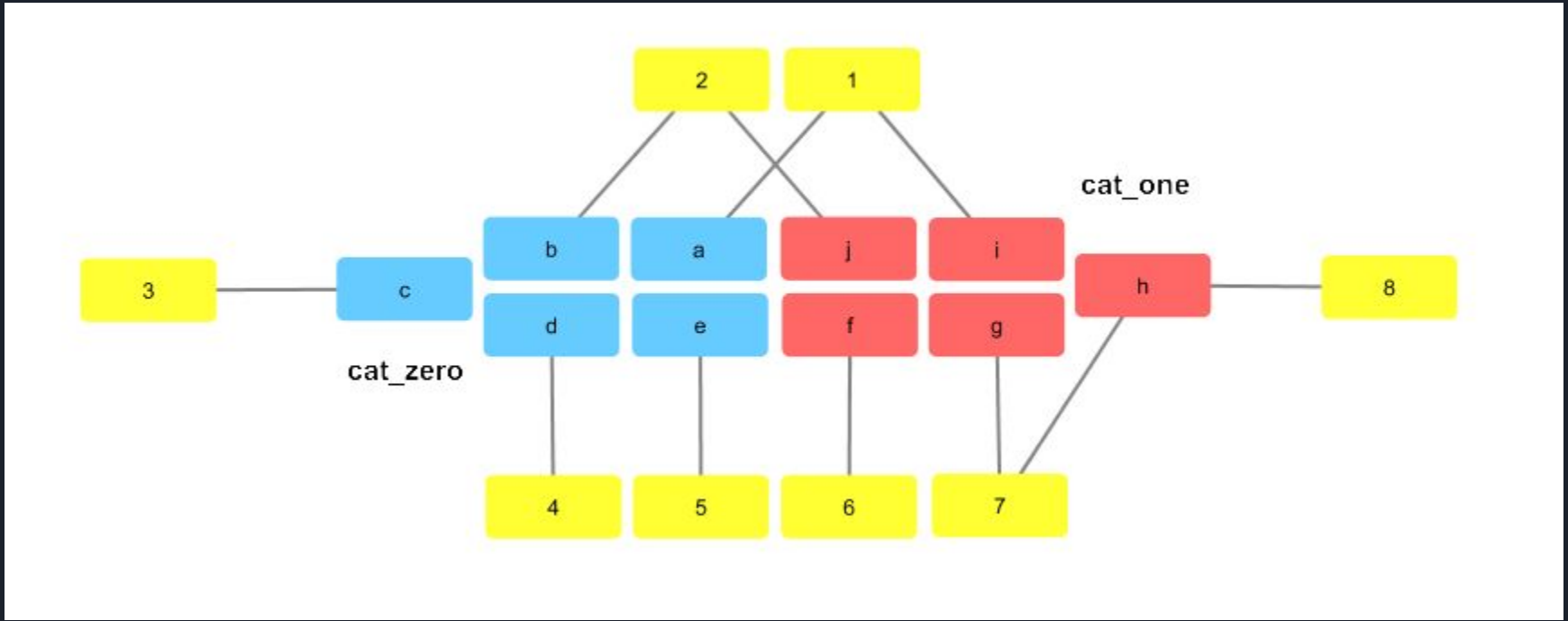
Hypothesis

In a network of samples labeled Obese or Lean, connected to nodes representing OTUs, when measuring the degree of OTUs associated with Obese nodes only, compared to those associated with Lean-only, will have a higher mean degree, which represents the higher diversity of OTUs outside the core microbiome.

Generally:

- In a network of samples and OTUs, a difference of the mean degree of OTUs associated with a boolean category is representative of a deviation from the core microbiome.

Simple Test Case:



The Script : Usage

network_analysis.py

-node {PATH to NODE FILE}

-edge {PATH to EDGE FILE}

[-o {PATH to OUTPUT DIRECTORY}]

-f {FEATURE COLUMN for comparison}

-c {CATEEGORY of FEATURE} {CATEGORY of FEATURE}

[-n {N_ITERATIONS for Monte Carlo Simulation}]

Under the Hood

Node File

node_nan	node_disp	ntype	degree	weighted	a_1	a_2	a_3	a_4	a_5	a_6	a_7	a_8	a_9	a_10	a_11	feature	a_12	a_13	a_14
1	otu_node	18	2	kpcofg	otu	otu	otu	otu	otu	otu	otu	otu	otu	otu	otu	otu	otu	otu	otu
2	otu_node	0	2	kpcofg	otu	otu	otu	otu	otu	otu	otu	otu	otu	otu	otu	otu	otu	otu	otu
3	otu_node	3	1	kpcofg	otu	otu	otu	otu	otu	otu	otu	otu	otu	otu	otu	otu	otu	otu	otu
4	otu_node	1	1	kpcofg	otu	otu	otu	otu	otu	otu	otu	otu	otu	otu	otu	otu	otu	otu	otu
5	otu_node	4	1	kpcofg	otu	otu	otu	otu	otu	otu	otu	otu	otu	otu	otu	otu	otu	otu	otu
6	otu_node	1	1	kpcofg	otu	otu	otu	otu	otu	otu	otu	otu	otu	otu	otu	otu	otu	otu	otu
7	otu_node	3	2	kpcofg	otu	otu	otu	otu	otu	otu	otu	otu	otu	otu	otu	otu	otu	otu	otu
8	otu_node	3	1	kpcofg	otu	otu	otu	otu	otu	otu	otu	otu	otu	otu	otu	otu	otu	otu	otu
a	a	user_node	1	1000	gibberish	nonsense	arbitrary	meaningle	unnecessa	number	zero	name	stuff	value	somethin	cat_zero	goods	services	pickles
b	b	user_node	1	1000	gibberish	nonsense	arbitrary	meaningle	unnecessa	number	one	name	stuff	value	somethin	cat_zero	goods	services	pickles
c	c	user_node	1	1000	gibberish	nonsense	arbitrary	meaningle	unnecessa	number	zero	name	stuff	value	somethin	cat_zero	goods	services	pickles
d	d	user_node	1	1000	gibberish	nonsense	arbitrary	meaningle	unnecessa	number	one	name	stuff	value	somethin	cat_zero	goods	services	pickles
e	e	user_node	1	1000	gibberish	nonsense	arbitrary	meaningle	unnecessa	number	zero	name	stuff	value	somethin	cat_zero	goods	services	pickles
f	f	user_node	1	1000	gibberish	nonsense	arbitrary	meaningle	unnecessa	number	one	name	stuff	value	somethin	cat_one	goods	services	pickles
g	g	user_node	1	1000	gibberish	nonsense	arbitrary	meaningle	unnecessa	number	zero	name	stuff	value	somethin	cat_one	goods	services	pickles
h	h	user_node	2	1000	gibberish	nonsense	arbitrary	meaningle	unnecessa	number	one	name	stuff	value	somethin	cat_one	goods	services	pickles
i	i	user_node	1	1000	gibberish	nonsense	arbitrary	meaningle	unnecessa	number	zero	name	stuff	value	somethin	cat_one	goods	services	pickles
j	j	user_node	1	1000	gibberish	nonsense	arbitrary	meaningle	unnecessa	number	one	name	stuff	value	somethin	cat_one	goods	services	pickles

Under the Hood

Edge File

from	to	eweight	a_1	a_2	a_3	a_4	a_5	a_6	a_7	a_8	a_9	a_10	a_11	feature	a_12	a_13	a_14
a	1	1	gibberish	nonsense	arbitrary	meaningl	unnecessa	number	zero	name	stuff	value	somethin	cat_zero	goods	services	pickles
b	2	1	gibberish	nonsense	arbitrary	meaningl	unnecessa	number	one	name	stuff	value	somethin	cat_zero	goods	services	pickles
c	3	1	gibberish	nonsense	arbitrary	meaningl	unnecessa	number	zero	name	stuff	value	somethin	cat_zero	goods	services	pickles
d	4	1	gibberish	nonsense	arbitrary	meaningl	unnecessa	number	one	name	stuff	value	somethin	cat_zero	goods	services	pickles
e	5	1	gibberish	nonsense	arbitrary	meaningl	unnecessa	number	zero	name	stuff	value	somethin	cat_zero	goods	services	pickles
f	6	1	gibberish	nonsense	arbitrary	meaningl	unnecessa	number	one	name	stuff	value	somethin	cat_one	goods	services	pickles
g	7	1	gibberish	nonsense	arbitrary	meaningl	unnecessa	number	zero	name	stuff	value	somethin	cat_one	goods	services	pickles
h	8	1	gibberish	nonsense	arbitrary	meaningl	unnecessa	number	one	name	stuff	value	somethin	cat_one	goods	services	pickles
i	1	1	gibberish	nonsense	arbitrary	meaningl	unnecessa	number	zero	name	stuff	value	somethin	cat_one	goods	services	pickles
j	2	1	gibberish	nonsense	arbitrary	meaningl	unnecessa	number	one	name	stuff	value	somethin	cat_one	goods	services	pickles
h	7	1	gibberish	nonsense	arbitrary	meaningl	unnecessa	number	one	name	stuff	value	somethin	cat_one	goods	services	pickles

Under the Hood

Join, and separate

from	to	feature	degree
a	1	cat_zero	2
i	1	cat_one	2
b	2	cat_zero	2
j	2	cat_one	2
c	3	cat_zero	1
d	4	cat_zero	1
e	5	cat_zero	1
f	6	cat_one	1
g	7	cat_one	2
h	7	cat_one	2
h	8	cat_one	1

- df_union

from	to	feature	degree
a	1	cat_zero	2
b	2	cat_zero	2
c	3	cat_zero	1
d	4	cat_zero	1
e	5	cat_zero	1

- cat_zero_table

from	to	feature	degree
i	1	cat_one	2
j	2	cat_one	2
f	6	cat_one	1
g	7	cat_one	2
h	7	cat_one	2
h	8	cat_one	1

- cat_one_table

Under the Hood

Further Separation

from	to	feature	degree
a	1	cat_zero	2
i	1	cat_one	2
b	2	cat_zero	2
j	2	cat_one	2
c	3	cat_zero	1
d	4	cat_zero	1
e	5	cat_zero	1
f	6	cat_one	1
g	7	cat_one	2
h	7	cat_one	2
h	8	cat_one	1

- df_union

from	to	feature	degree
c	3	cat_zero	1
d	4	cat_zero	1
e	5	cat_zero	1

- otu_zero_only

from	to	feature	degree
h	6	cat_one	1
i	7	cat_one	1
j	8	cat_one	1

- otu_one_only

from	to	feature	degree
a	1	cat_zero	2
b	2	cat_zero	2
f	1	cat_one	2
g	2	cat_one	2

- otu_both

Under the Hood

Reduction

otu_zero_only

from	to	feature	degree
c	3	cat_zero	1
d	4	cat_zero	1
e	5	cat_zero	1

from	to	feature	degree
c	3	cat_zero	1
d	4	cat_zero	1
e	5	cat_zero	1

otu_zero_only

otu_one_only

from	to	feature	degree
h	6	cat_one	1
i	7	cat_one	1
j	8	cat_one	1

from	to	feature	degree
h	6	cat_one	1
i	7	cat_one	1
j	8	cat_one	1

otu_one_only

otu_both

from	to	feature	degree
a	1	cat_zero	2
b	2	cat_zero	2
f	1	cat_one	2
g	2	cat_one	2

from	to	feature	degree
a	1	cat_zero	2
b	2	cat_zero	2

otu_both

Under the Hood

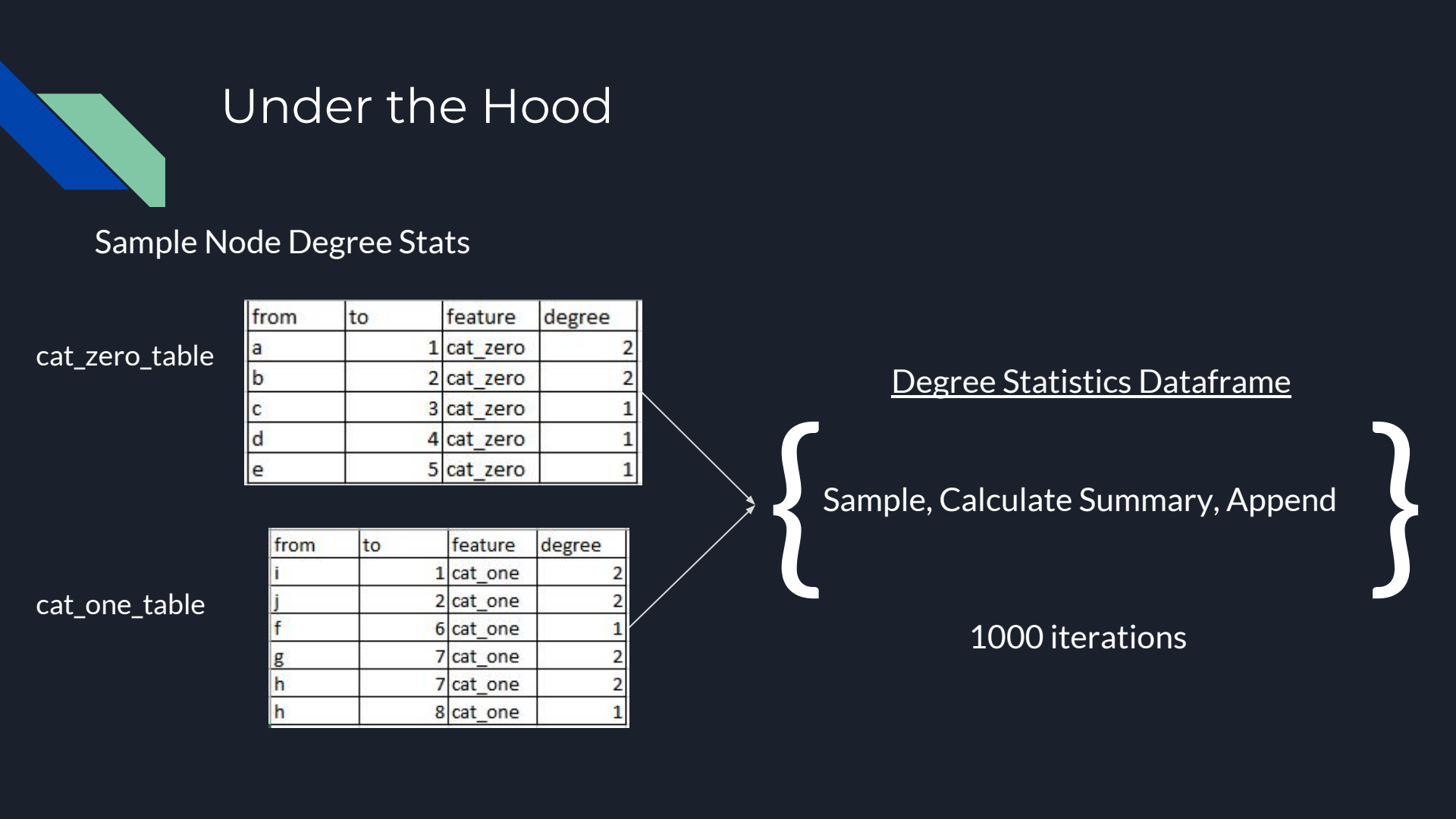
Sample Node Degree Stats

cat_zero_table

from	to	feature	degree
a	1	cat_zero	2
b	2	cat_zero	2
c	3	cat_zero	1
d	4	cat_zero	1
e	5	cat_zero	1

cat_one_table

from	to	feature	degree
i	1	cat_one	2
j	2	cat_one	2
f	6	cat_one	1
g	7	cat_one	2
h	7	cat_one	2
h	8	cat_one	1



Under the Hood

OTU Node Degree Stats

otu_zero_only

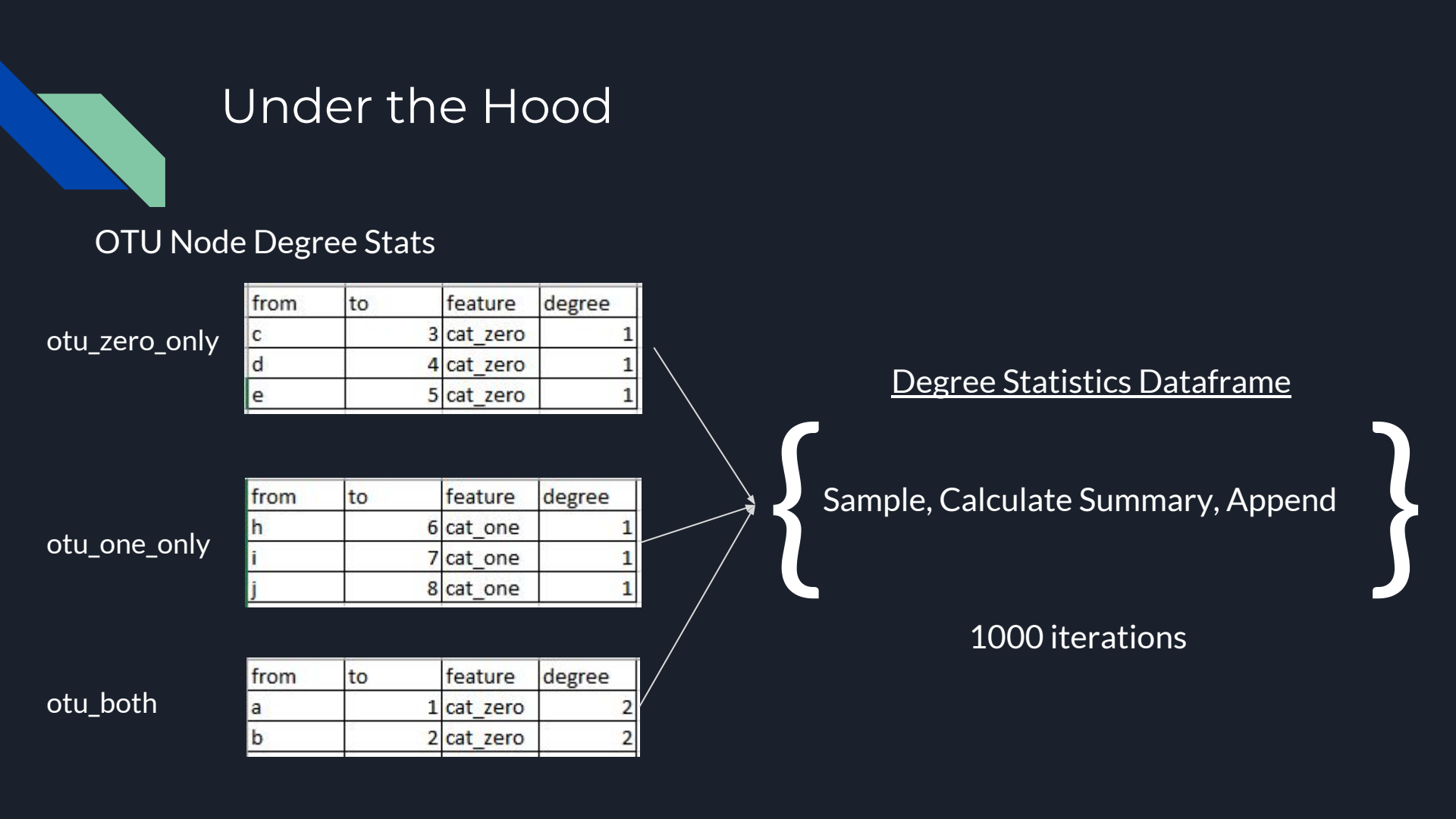
from	to	feature	degree
c	3	cat_zero	1
d	4	cat_zero	1
e	5	cat_zero	1

otu_one_only

from	to	feature	degree
h	6	cat_one	1
i	7	cat_one	1
j	8	cat_one	1

otu_both

from	to	feature	degree
a	1	cat_zero	2
b	2	cat_zero	2



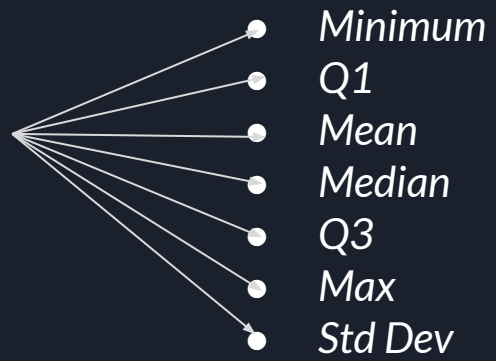
Under the Hood

Degree Statistics Dataframe

Sample, Calculate Summary, Append

1000 iterations

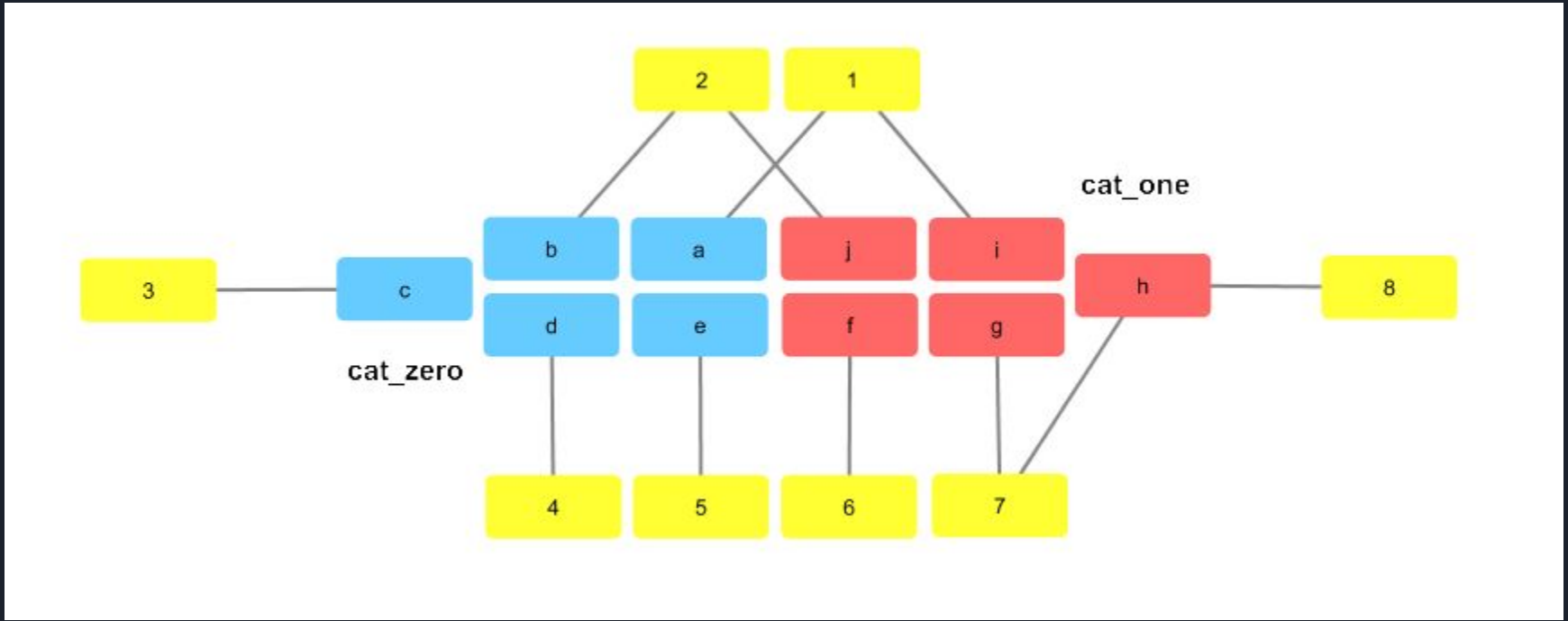
Output Statistics



Minimum
Q1
Mean
Median
Q3
Max
Std Dev

column.mean()

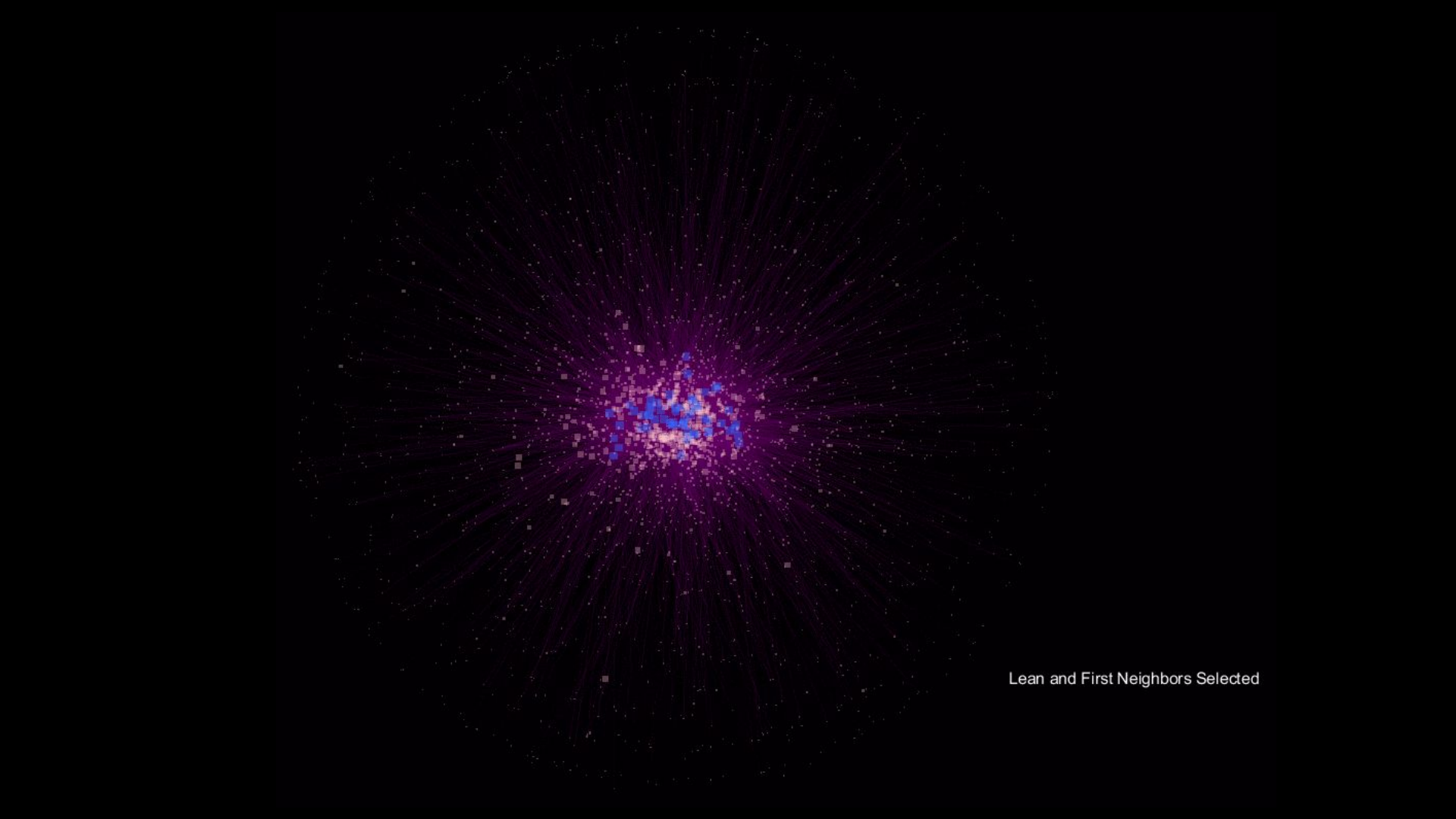
Test Case Network



Under the Hood

Output statistics

	Minimum	Q1	Mean	Median	Q3	Max	Std Dev
cat_zero	1.0	1.0	1.396	1.095	1.967	2.0	0.489
cat_one	1.0	1.155	1.669	1.981	2.0	2.0	0.47
otu_zero	1.0	1.0	1.0	1.0	1.0	1.0	0.0
otu_one	1.0	1.0	1.331	1.015	1.841	2.0	0.47
otu_both	2.0	2.0	2.0	2.0	2.0	2.0	0.0

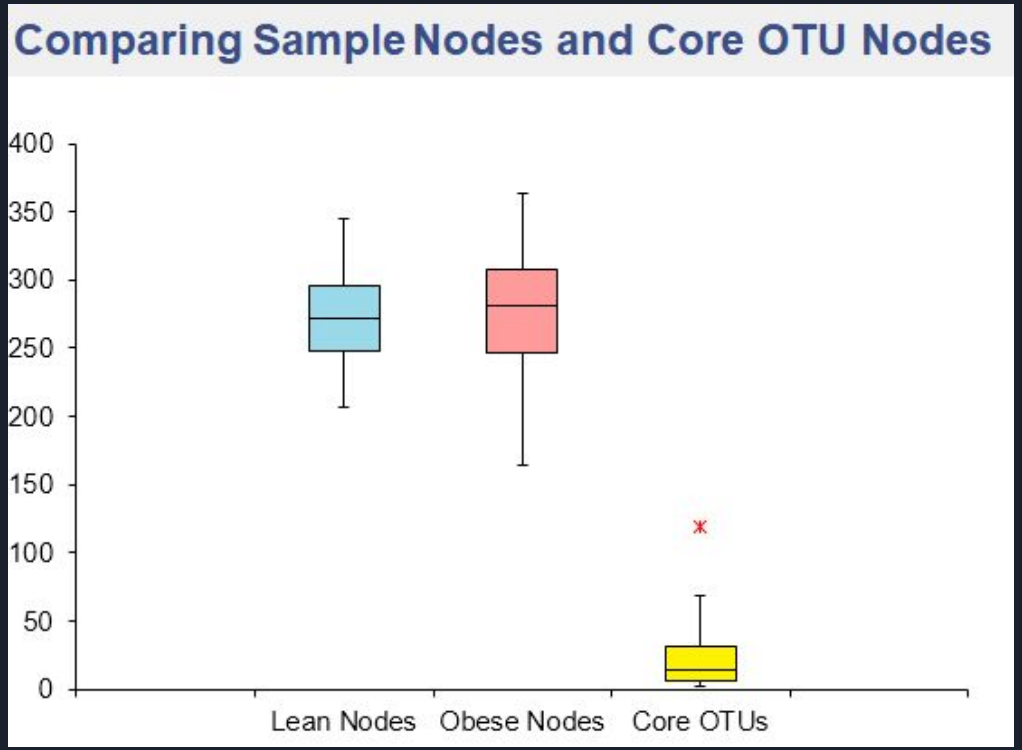

Lean and First Neighbors Selected

Obese vs. Lean Output

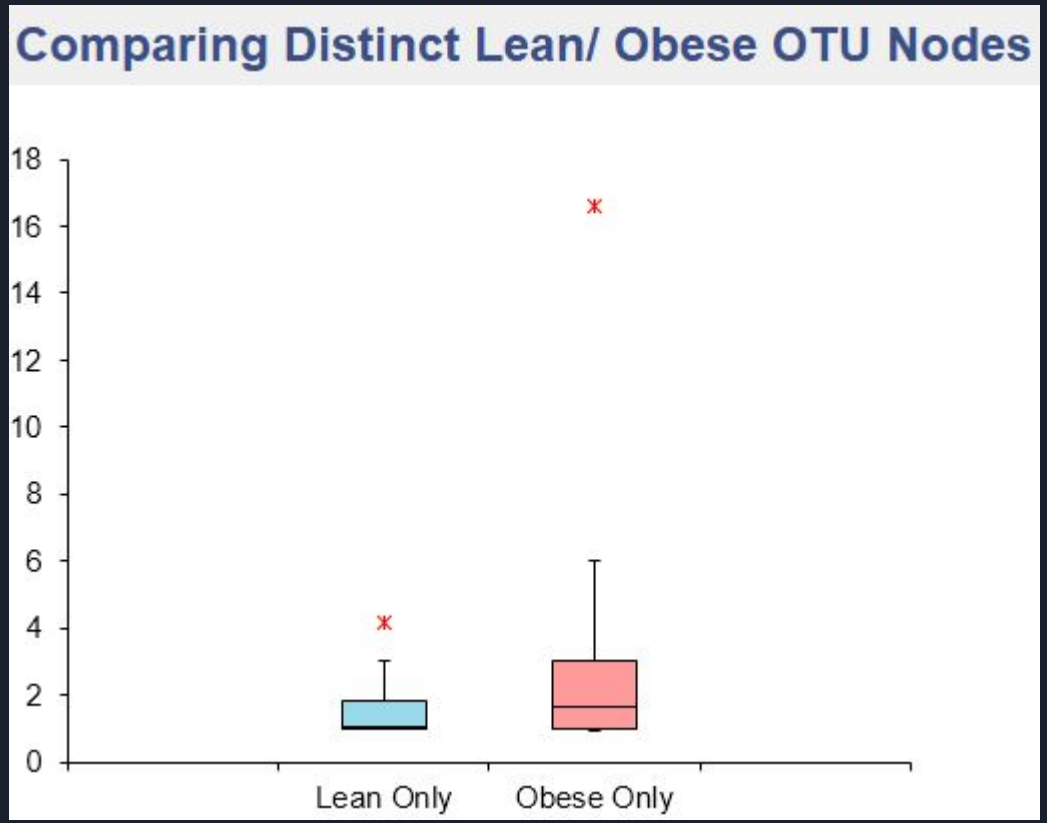
Output statistics

	Minimum	Q1	Mean	Median	Q3	Max	Std Dev
Lean	206.103	248.114	271.906	271.788	296.252	344.526	35.574
Obese	163.965	246.257	275.77	280.659	308.187	363.067	46.73
otu_Lean	1.0	1.0	1.435	1.01	1.809	4.185	0.759
otu_Obese	1.0	1.003	2.732	1.644	3.016	16.614	3.154
otu_Both	2.067	6.385	24.117	14.049	31.4	119.475	26.959

Obese nodes have a wider distribution of OTU connections, and thus a more diverse gut microbiome.



OTUs associated with Lean-only nodes have a low mean degree when compared to those of Obese-only nodes.



Conclusion

- The degree of OTUs associated with Obese-only nodes have a higher degree when compared to those of Lean-only nodes. This supports the notion of deviations from the core microbiome being associated with physiological states; in this case, Obesity.

github