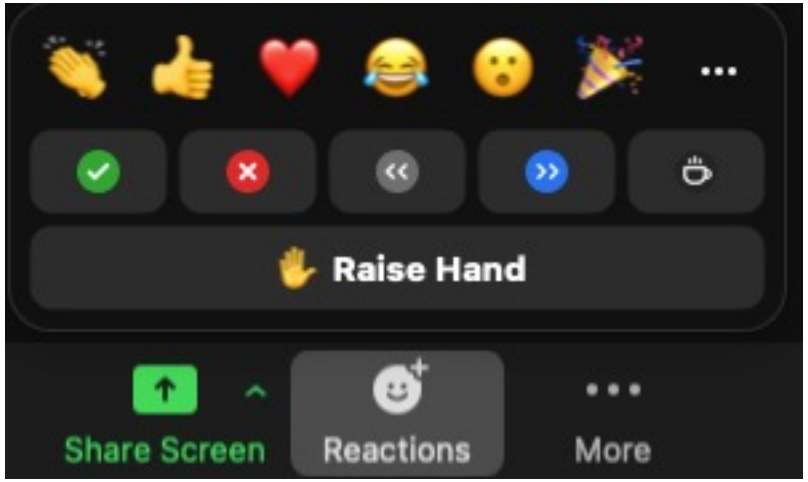
Workshop 2: Network analysis in conservation

Quantitative skills workshop: outline & logistics

- Target time of ~2 hours
- 10 minute breaks after 55 minutes
- Interrupt when you need to, can also "raise hand"
- Zoom will be recorded

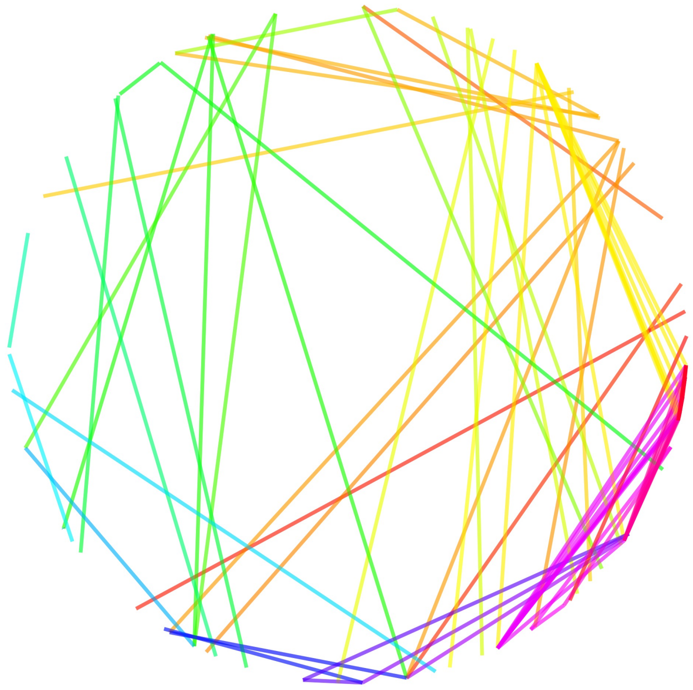


GitHub:
https://github.com/MikeBode/SAEF_quantitative_skills

Google CoLaborate:
https://research.google.com/colaboratory/
(need a Gmail account)

Network & graph theory

- Branch of discrete mathematics
- Objects (vertices) joined by relationships (edges)
- These mathematical structures can be represented as graphs or matrices
- Extremely flexible framework. Nodes can be individuals, patches, genes, communities, populations. Edges can be social, spatial, correlations, demographic, disease spread, pollination.

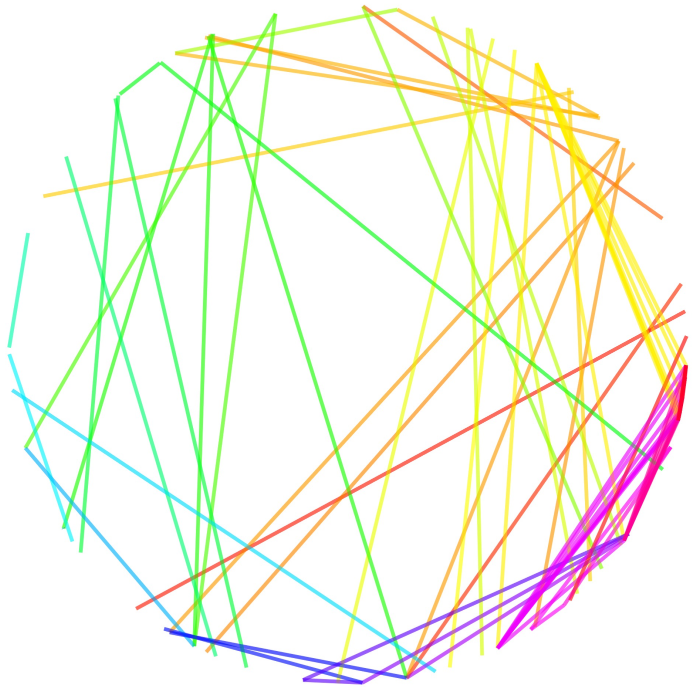


```
R =

  10×10 logical array

   0   1   1   1   0   1   0   0   1   0
   0   1   0   1   0   0   1   1   1   1
   1   0   0   1   1   0   0   0   1   1
   0   1   0   1   1   0   0   0   1   0
   1   1   0   0   0   0   0   0   0   1
   1   1   0   1   1   1   1   1   1   0
   1   1   0   0   1   0   0   1   0   1
   0   0   1   1   0   1   1   1   0   1
   0   0   0   1   1   0   0   0   1   0
   0   1   1   0   0   0   1   1   0   1
```

Network & graph theory

1. Network metrics (prac)
2. Statistical network analysis (theory)
3. Modularity analysis (prac)
4. Dynamical systems on networks (theory)

Network theory is elegant, powerful, and frequently abused.



```
R =

  10×10 logical array

   0   1   1   1   0   1   0   0   1   0
   0   1   0   1   0   0   1   1   1   1
   1   0   0   1   1   0   0   0   1   1
   0   1   0   1   1   0   0   0   1   0
   1   1   0   0   0   0   0   0   0   1
   1   1   0   1   1   1   1   1   1   0
   1   1   0   0   1   0   0   1   0   1
   0   0   1   1   0   1   1   1   0   1
   0   0   0   1   1   0   0   0   1   0
   0   1   1   0   0   0   1   1   0   1
```

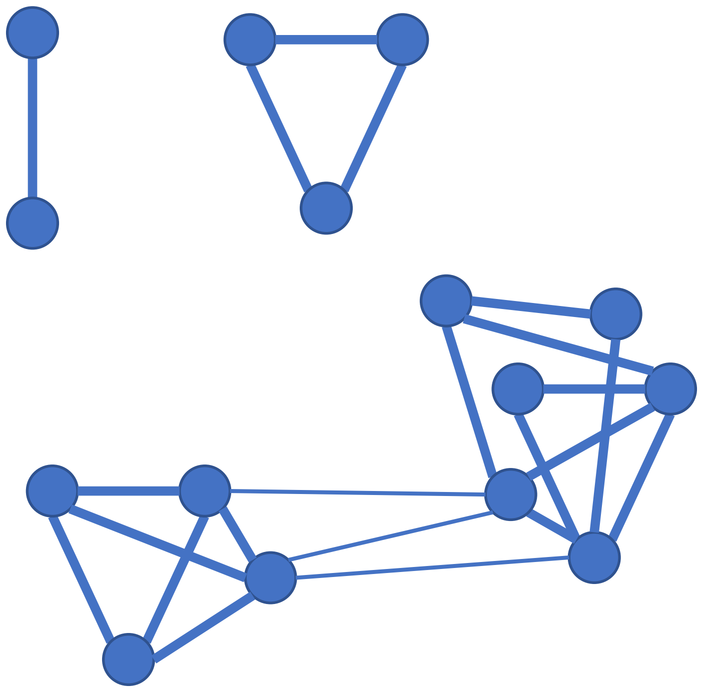
Calculating network metrics

Networks are complex objects.

- N^2 direct connections among N objects
- N^3 tertiary relationships (etcetera)
- Very large numbers of meaningful partitions (e.g., groups, factions, coalitions) of heterogeneous size
- Weighted & directed
- Covariates and categorical differences between nodes & edges.

We want them to have simple meaning

```
R =

  10×10 logical array

   0   1   1   1   0   1   0   0   1   0
   0   1   0   1   0   0   1   1   1   1
   1   0   0   1   1   0   0   0   1   1
   0   1   0   1   1   0   0   0   1   0
   1   1   0   0   0   0   0   0   0   1
   1   1   0   1   1   1   1   1   1   0
   1   1   0   0   1   0   0   1   0   1
   0   0   1   1   0   1   1   1   0   1
   0   0   0   1   1   0   0   0   1   0
   0   1   1   0   0   0   1   1   0   1
```



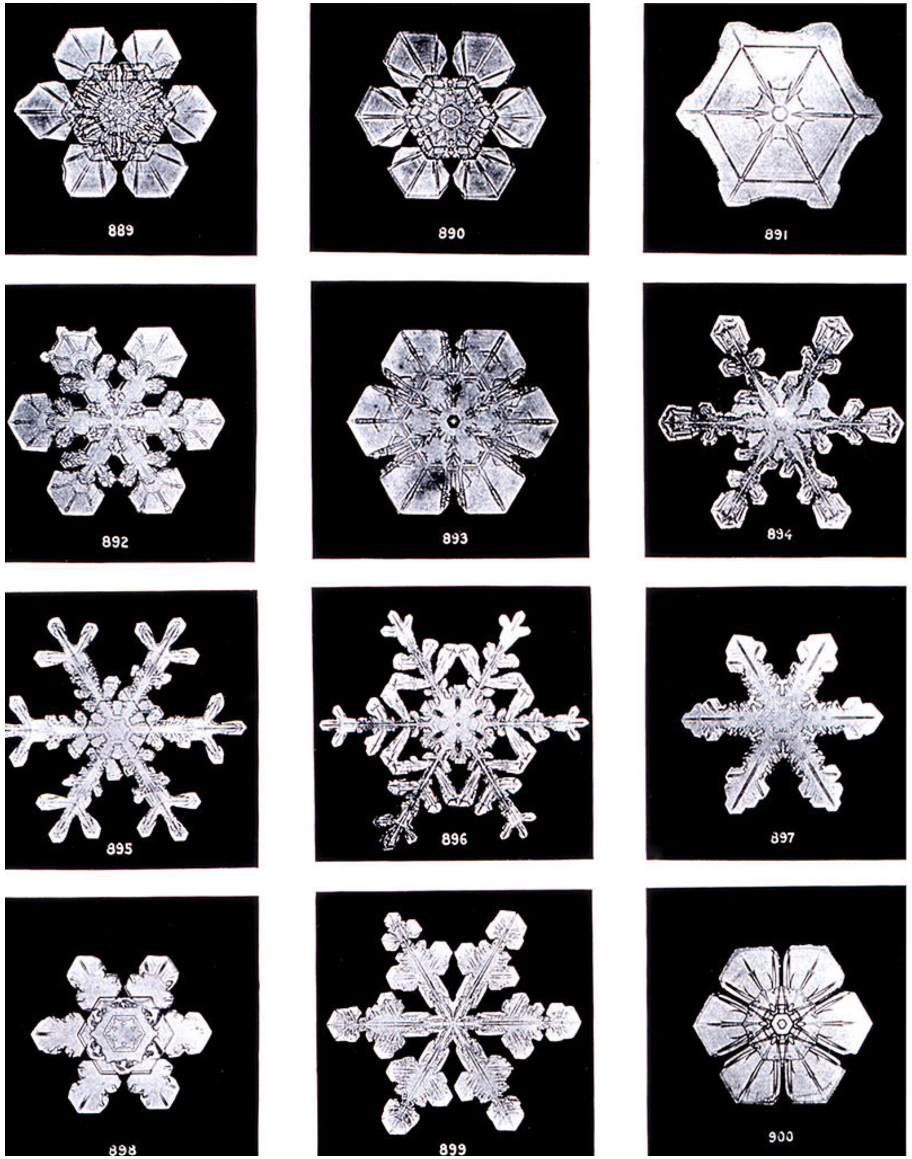
Calculating network metrics

Network metrics offer simple insights.

- Scalar properties of the network
- Local descriptions of regional structure
- Lower dimensional structure

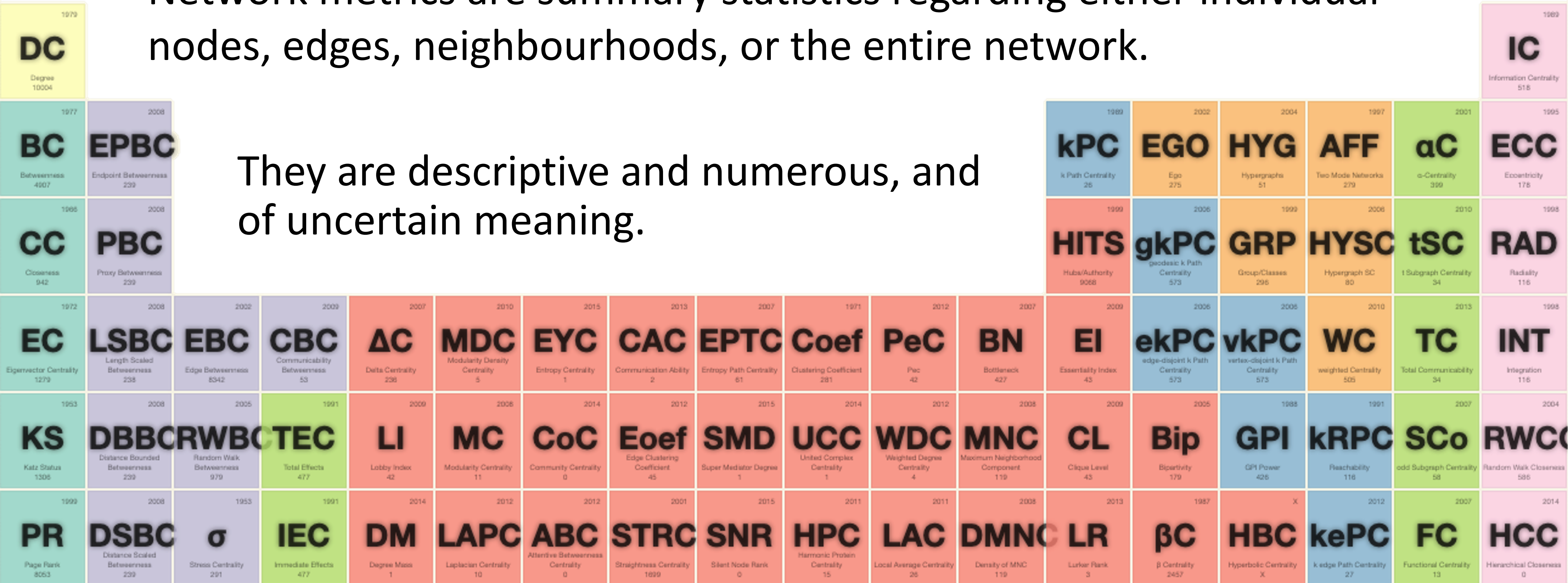
Describing complexity with simplicity requires either (1) emergence, or (2) acceptance of loss.

The number, variety, and non-independence of network metrics makes them scientifically fraught.

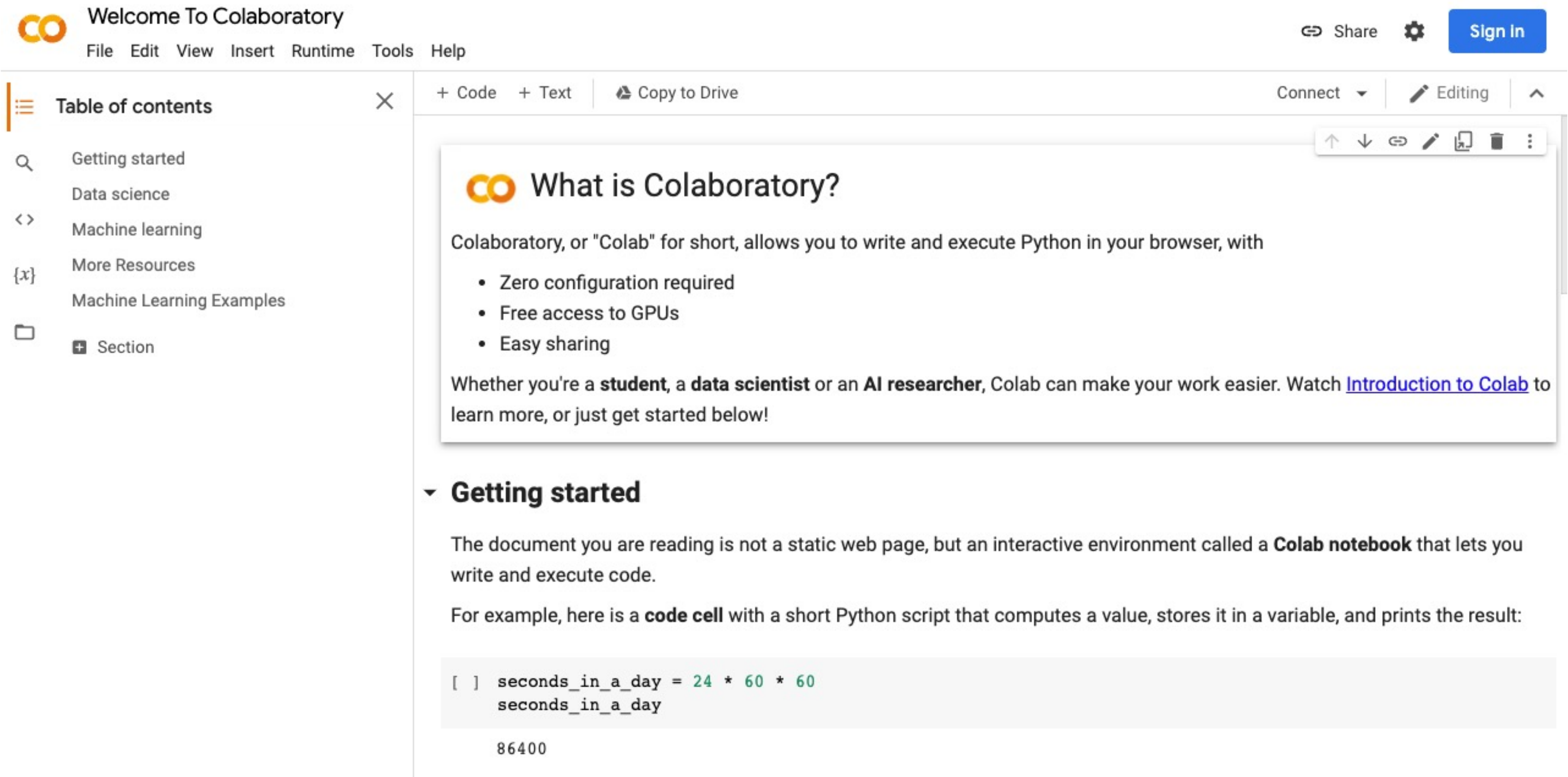


Network metrics are summary statistics regarding either individual nodes, edges, neighbourhoods, or the entire network.

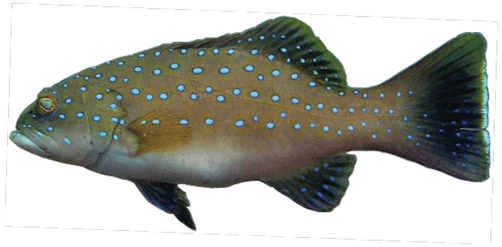
They are descriptive and numerous, and of uncertain meaning.



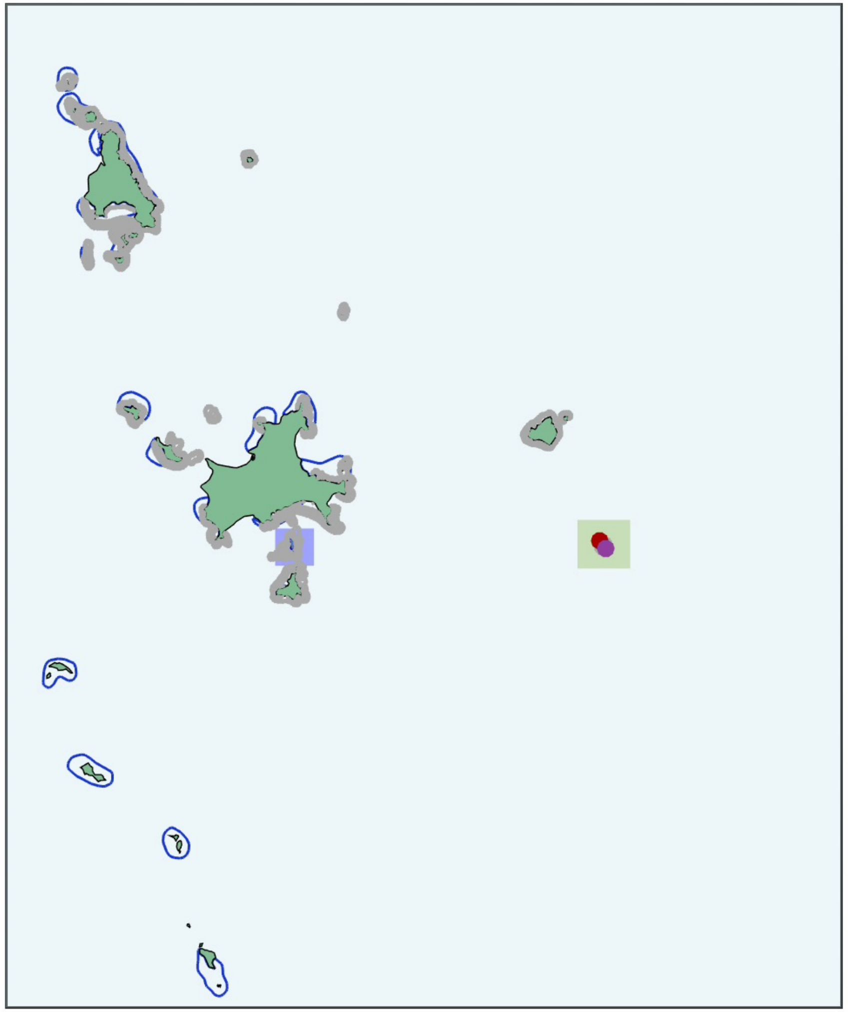
Practical task set 1: metrics in marine ecology



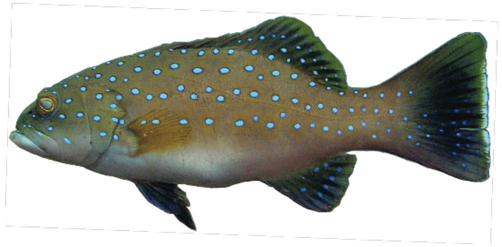
Larval dispersal networks



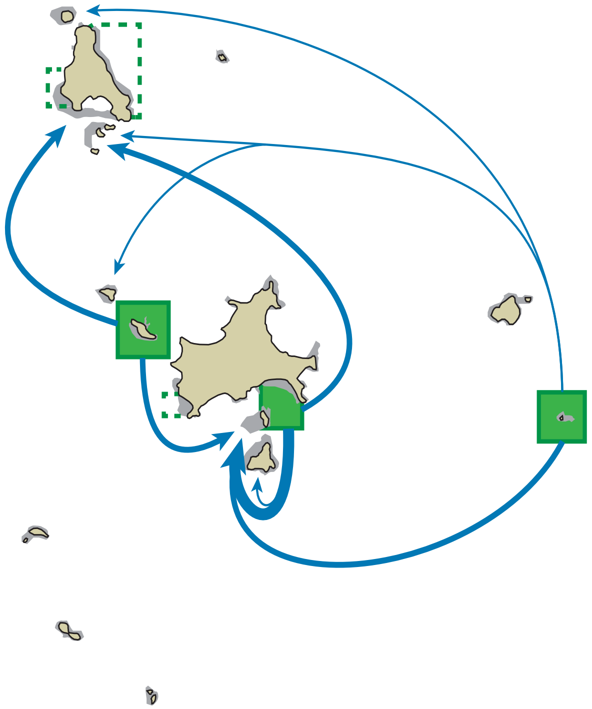
- Coral reef fishes generally spend their adult lives on a single reef
- Their offspring have an obligate pelagic larval dispersal phase that can last months
- The resulting demographic exchange is essential for understanding the ecology, evolution, and conservation of these species



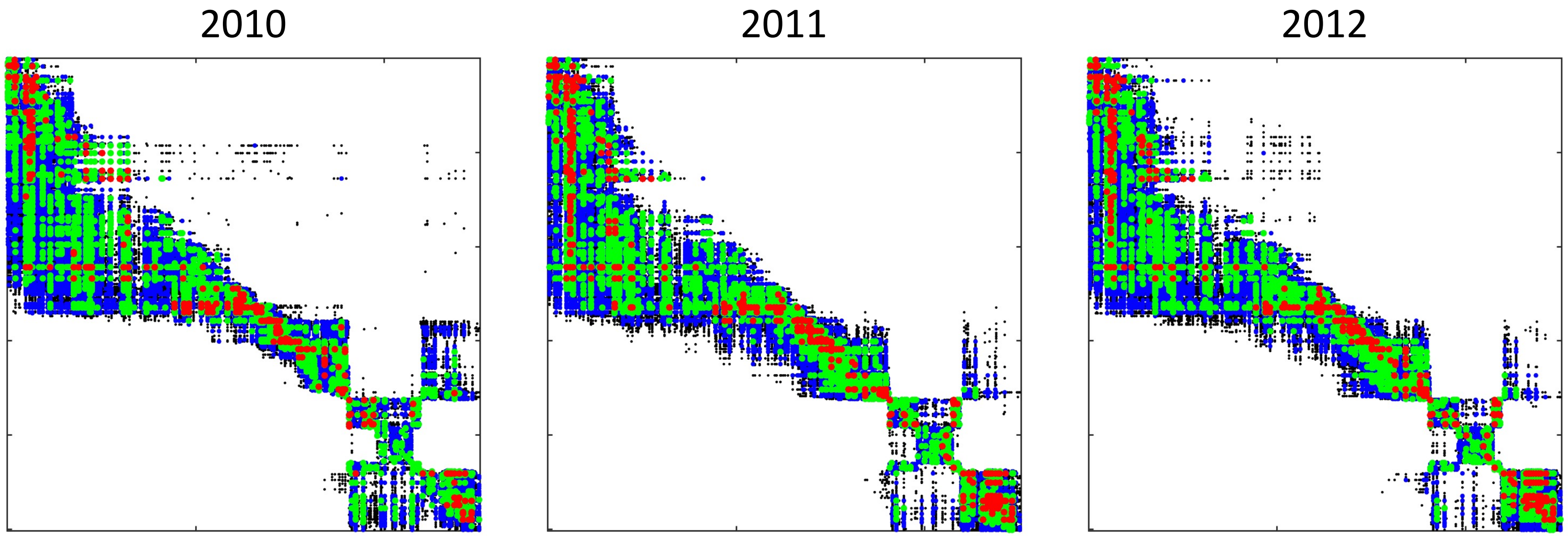
Larval dispersal networks

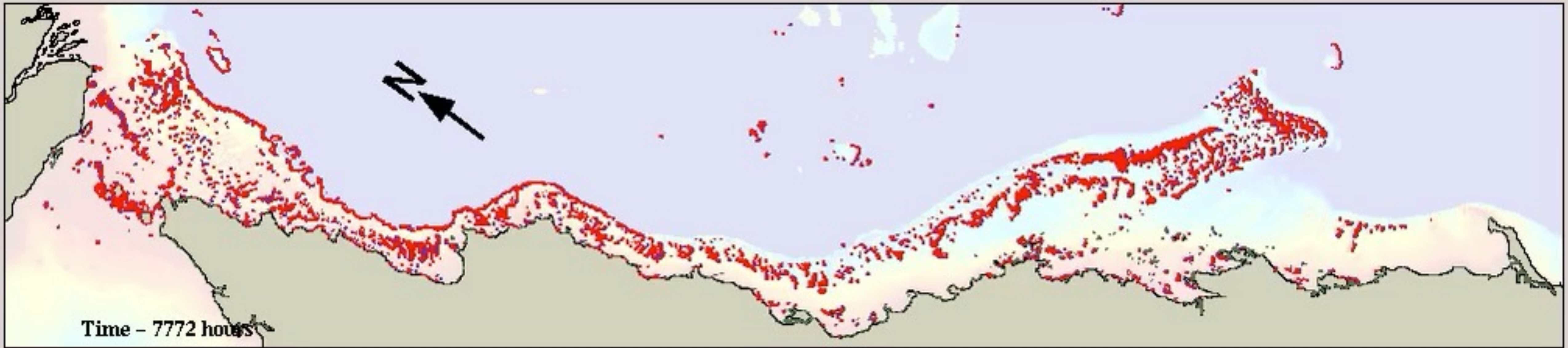


- The resulting pattern of exchange can be characterised as networks or graphs
- These networks are directed, weighted, stochastic graphs.



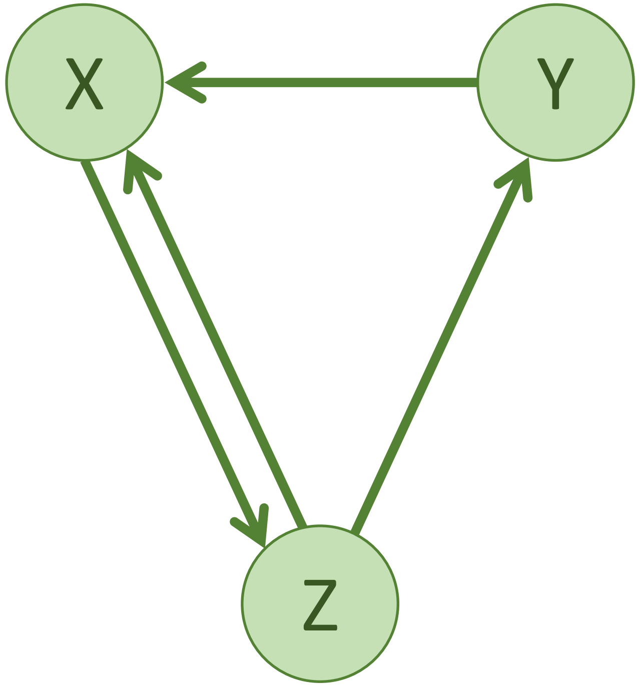
Can network theory metrics offer simple insights into these complex systems?





In- and out-degree distributions

- The in-degree of a node measures the number of vertices that enter a node
- The out-degree measures the number that exit a node



$D_{out}(X) = 1$ $D_{in}(X) = 2$

$D_{out}(Y) = 1$ $D_{in}(Y) = 1$

$D_{out}(Z) = 2$ $D_{in}(Z) = 1$

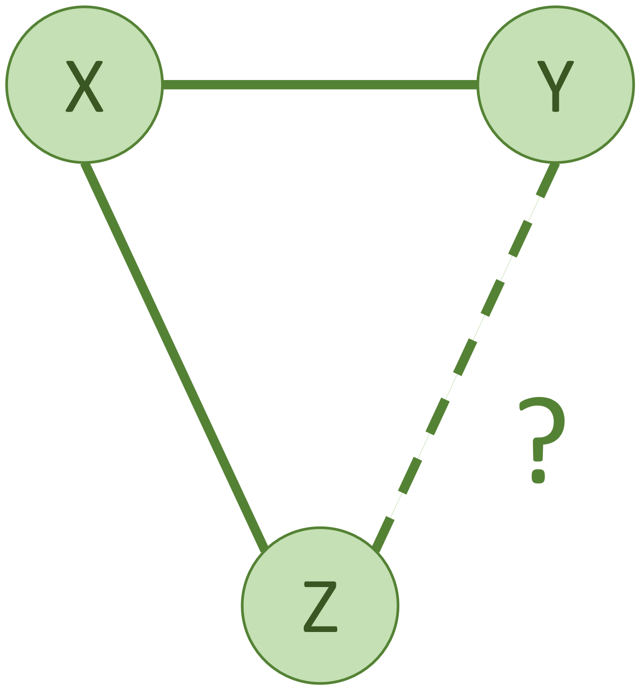
Note that $\sum_i D_{\text{out}}(i) = \sum_i D_{\text{in}}(i)$

Clustering coefficient

Collective dynamics of ‘small-world’ networks

Duncan J. Watts* & Steven H. Strogatz

- Node X is connected to node Y
- Node X is also connected to node Y
- What is the probability that nodes Y and Z are connected?
- This would make X, Y, and Z a “clique”



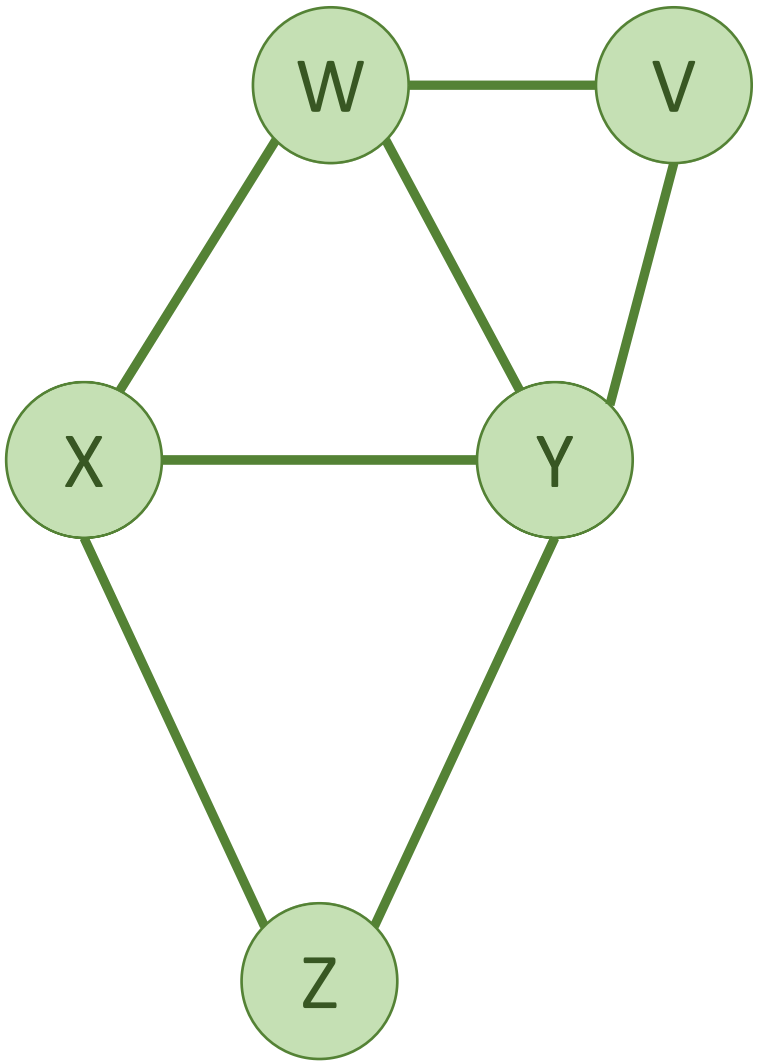
Clustering coefficient

The clustering coefficient of a node X is the proportion of potential triangles among the neighbours of X that are realised.

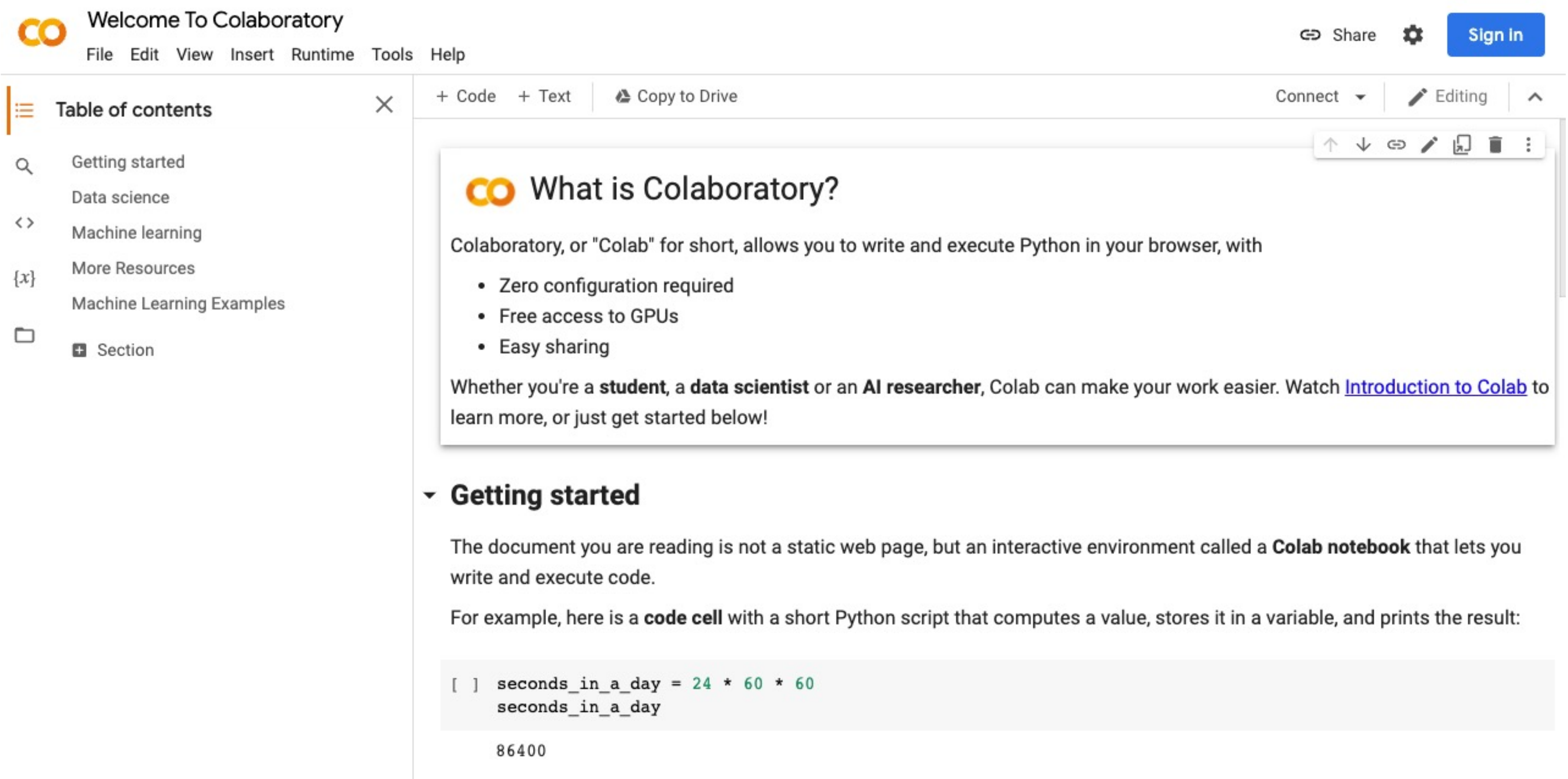
C(X) = 2/3　　C(Y) = 1/2

C(W) = 1　　C(Z) = 1

C($\mathcal{G}$) = 0.792

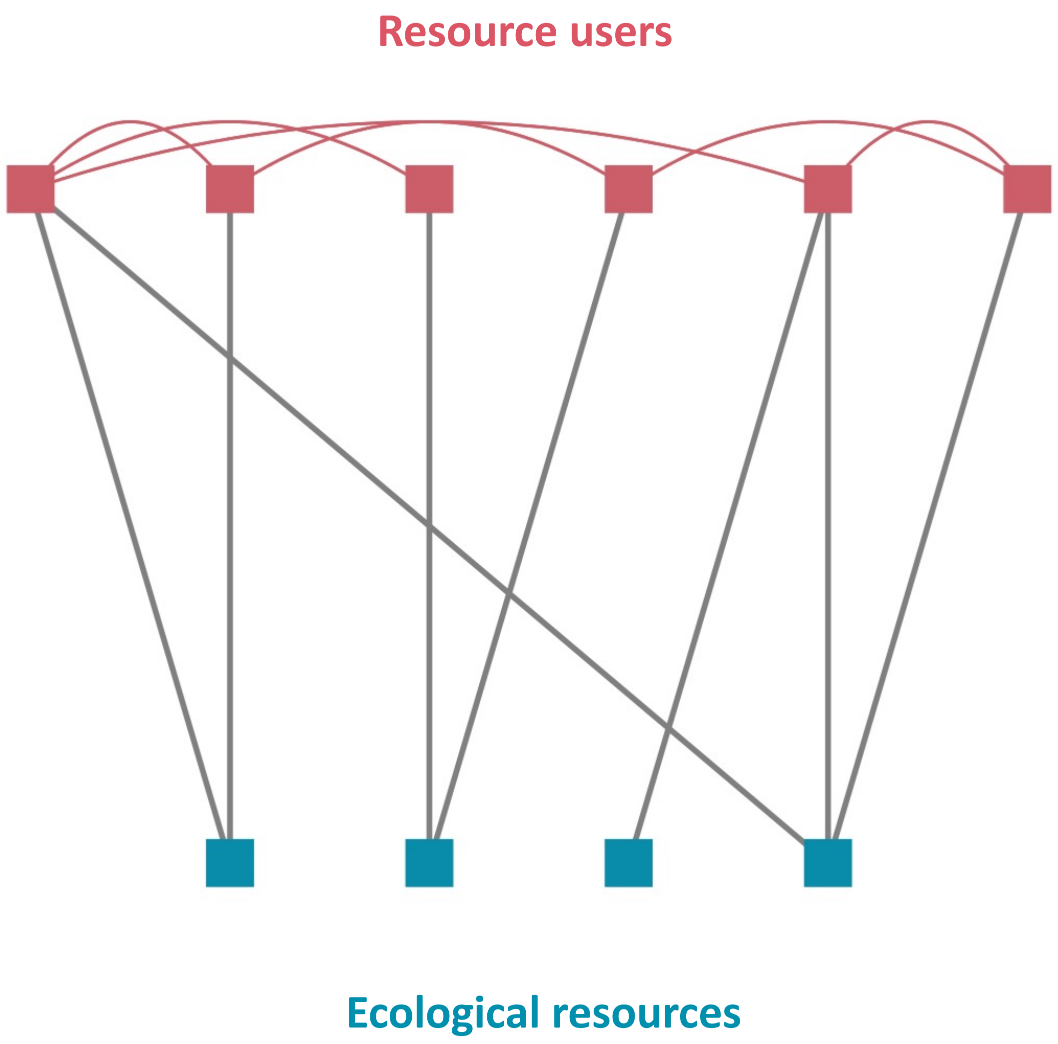


Practical task set 1: metrics in marine ecology



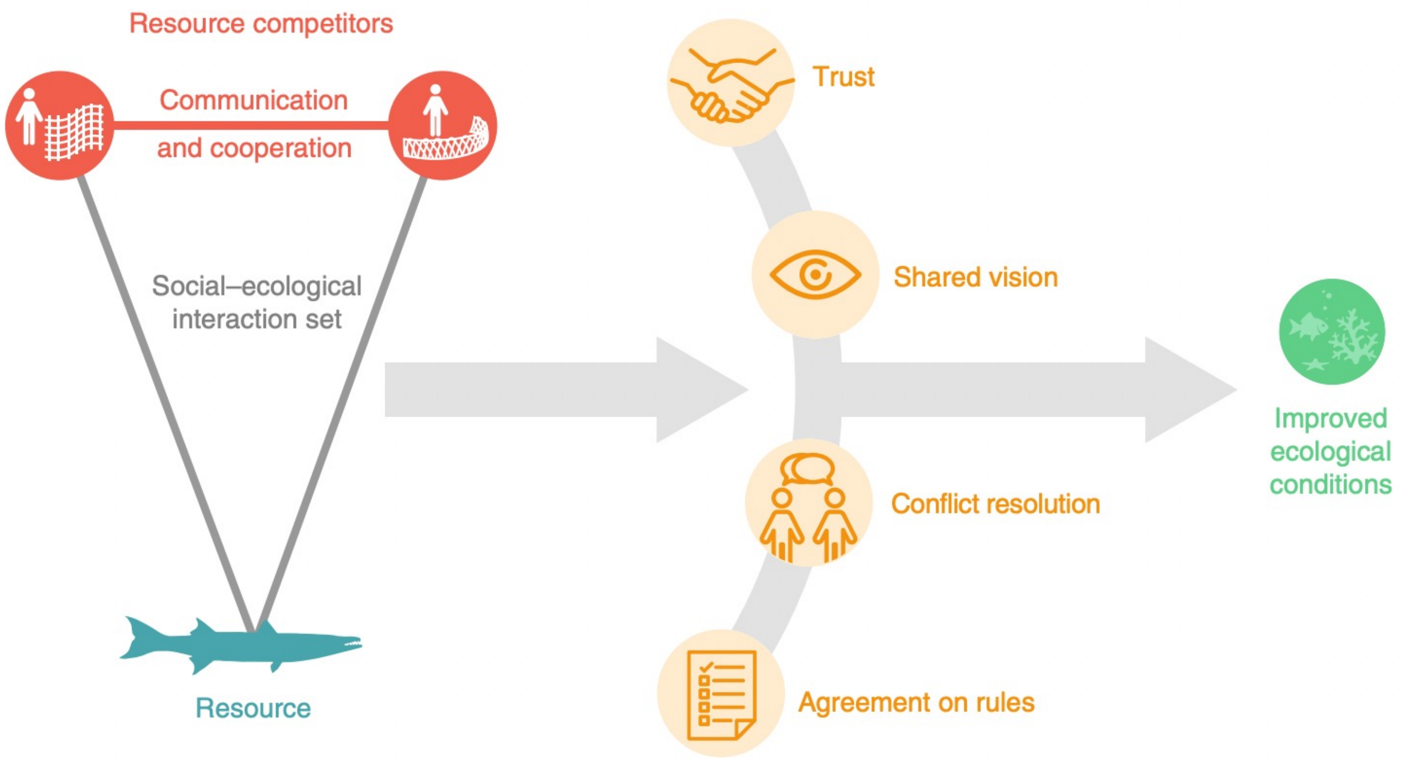
Statistical Social Network Analysis

- Bipartite socio-ecological networks contain two categories of object **social** and **ecological**.
- **Social** objects are connected to each other through **social** relationships.
- **Ecological** objects are connected to **social** objects through **socioecological** relationships
- **Ecological** objects are not connected to each other in this model



Statistical Social Network Analysis

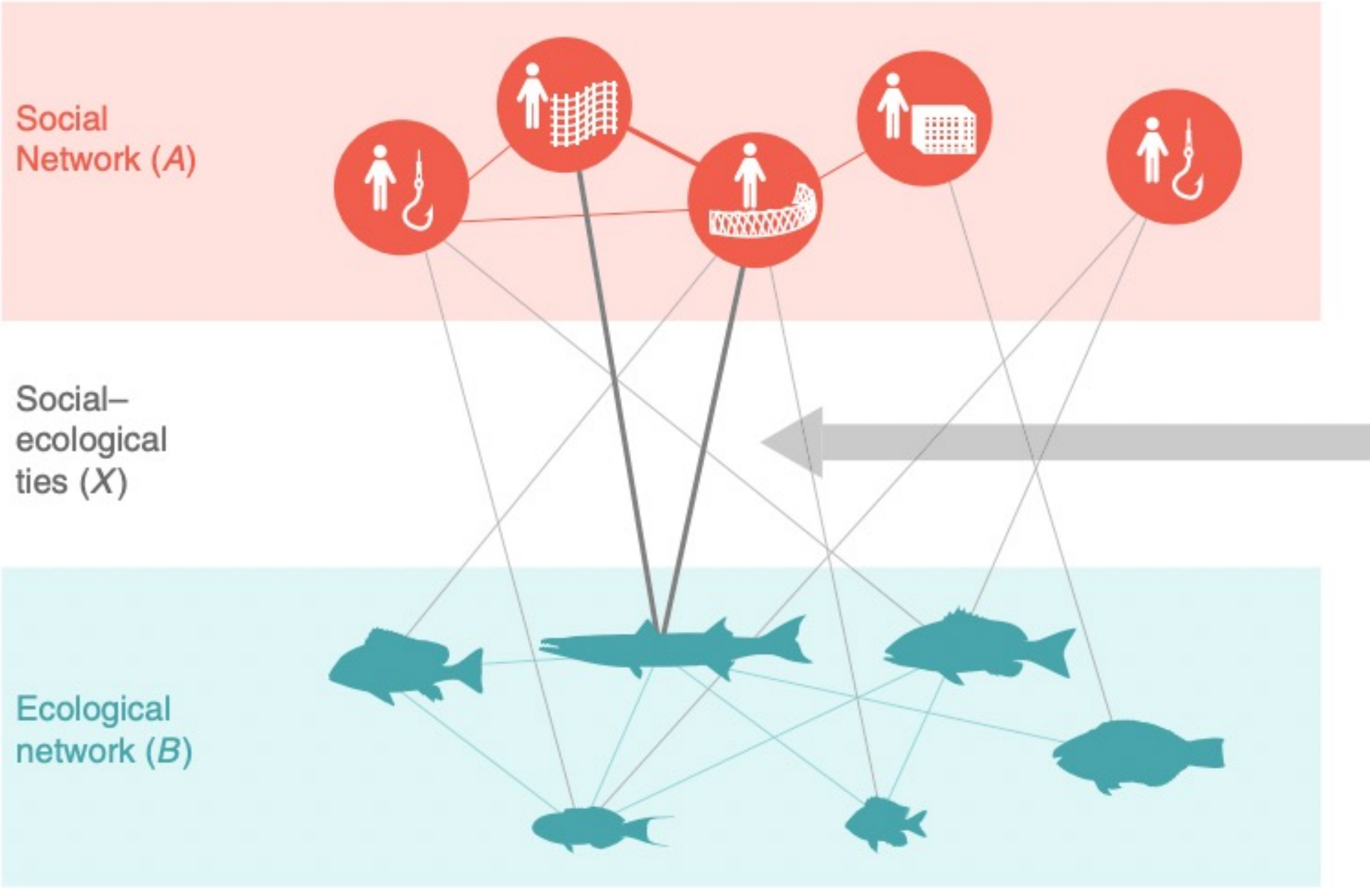
- In social network science, closure facilitates trust and learning among actors who share common (or interconnected) resources.
- This behaviour supports and reinforces norms, and facilitates learning and consensus.
- It is thought to be frequently observed in common-pool resource systems that are well-managed.



Barnes *et al*. (2019) Social-ecological alignment and ecological conditions in coral reefs. *Nature Communications*

Statistical Social Network Analysis

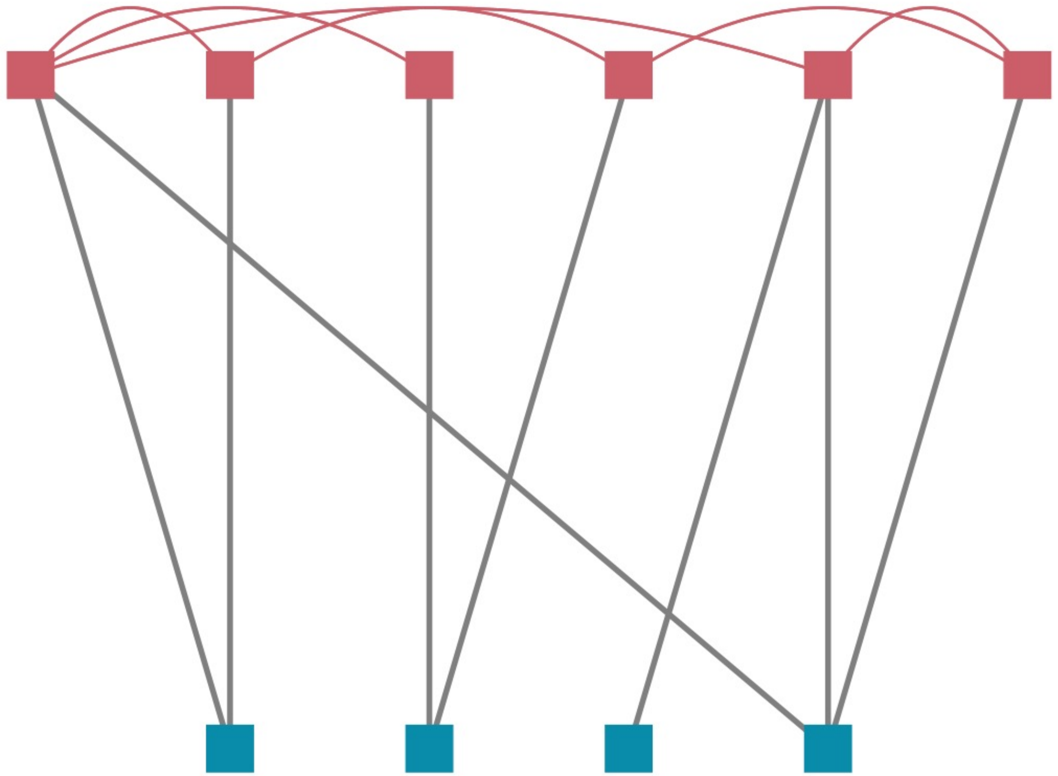
- Barnes *et al*. interviewed several hundred Kenyan fishers who exploit dozens of marine species.
- They search the network for closed socio-ecological triangles.
- They concluded that positive ecological conditions are associated with socio-ecological network closure.
- Specifically, they found statistically higher fish biomass (p=0.03) and higher functional richness (p<0.01).



Barnes *et al*. (2019) Social-ecological alignment and ecological conditions in coral reefs. *Nature Communications*

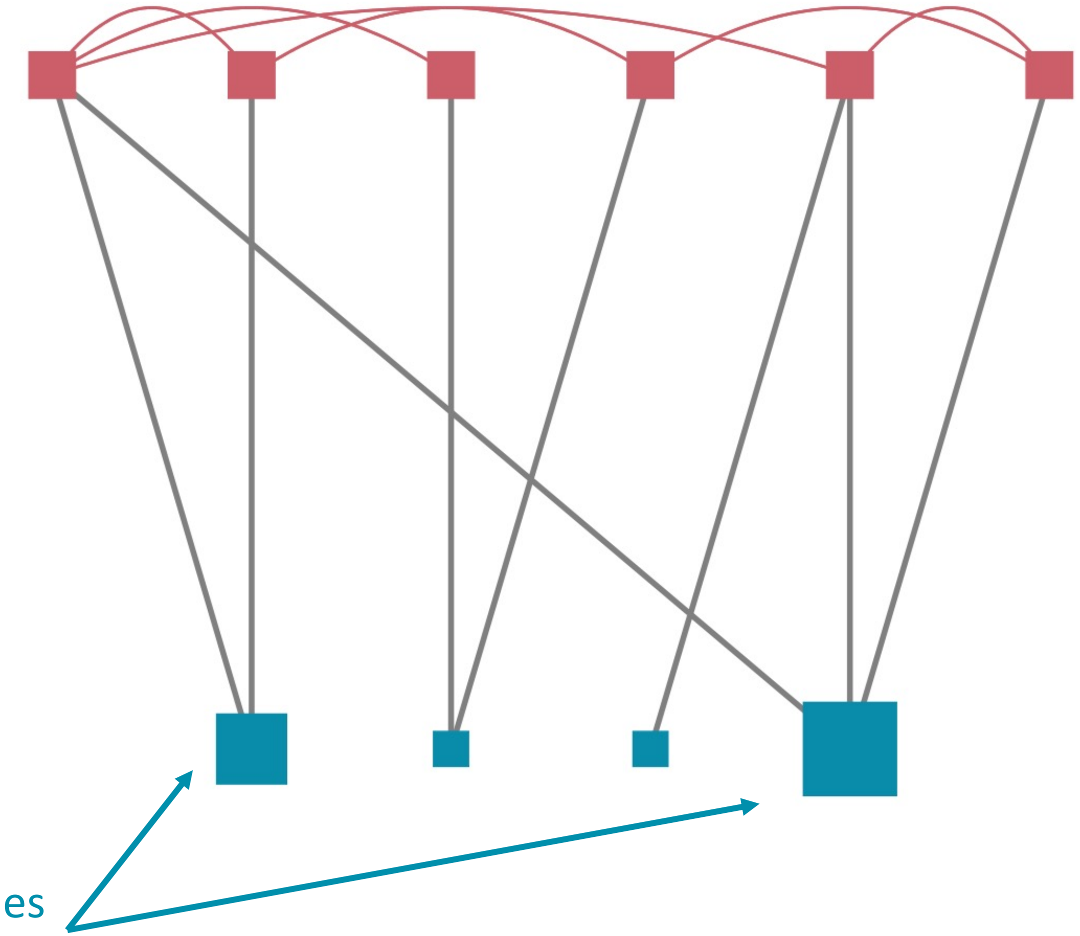
Statistical Social Network Analysis

- What analyses can allow us to draw statistical conclusions about how SE network structures determine ecological outcomes?



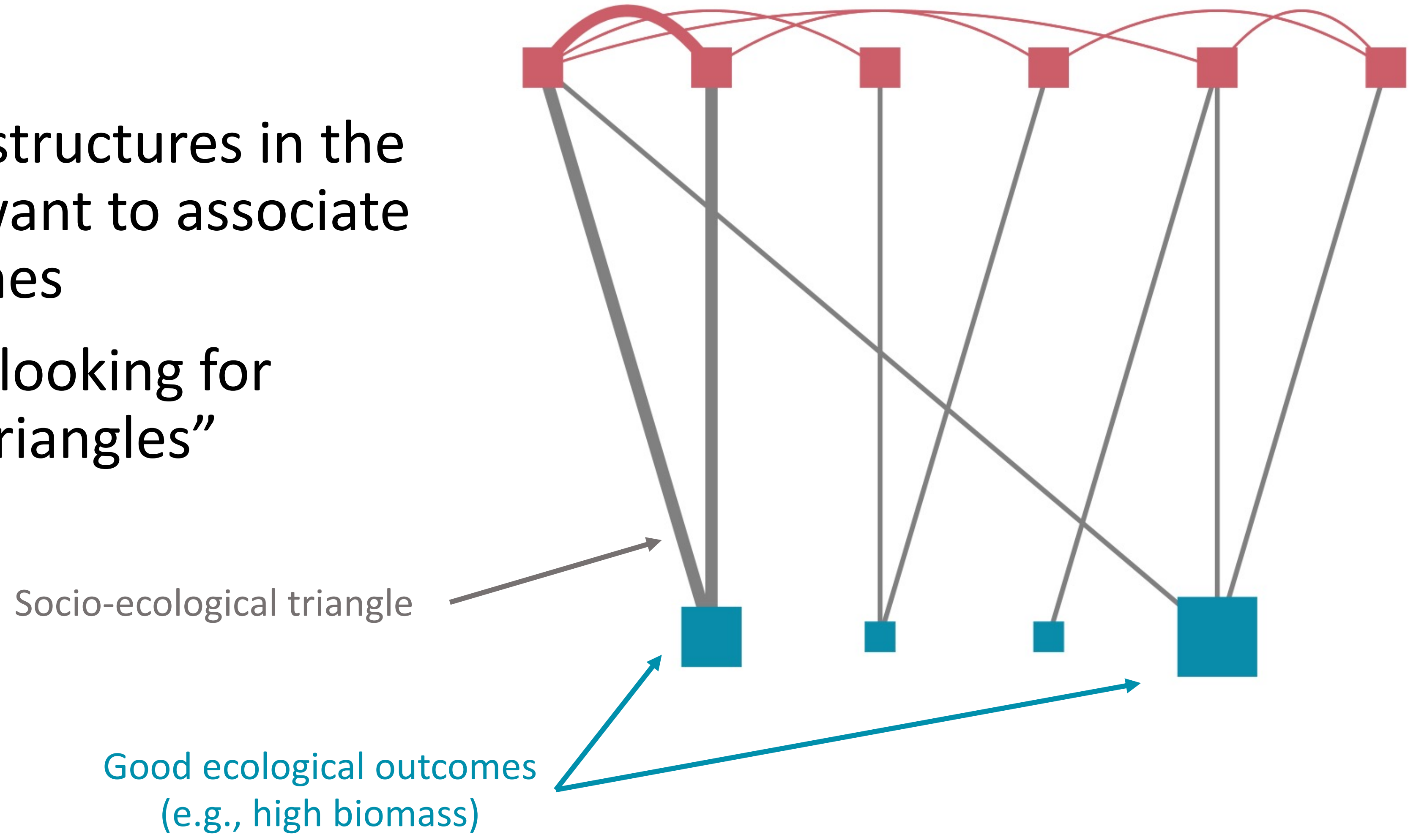
Statistical Social Network Analysis

- First, we measure the ecological outcomes of each ecological node in the network.



Statistical Social Network Analysis

- Then, we identify structures in the network that we want to associate with these outcomes
- In this case, we're looking for "socio-ecological triangles"



Statistical Social Network Analysis

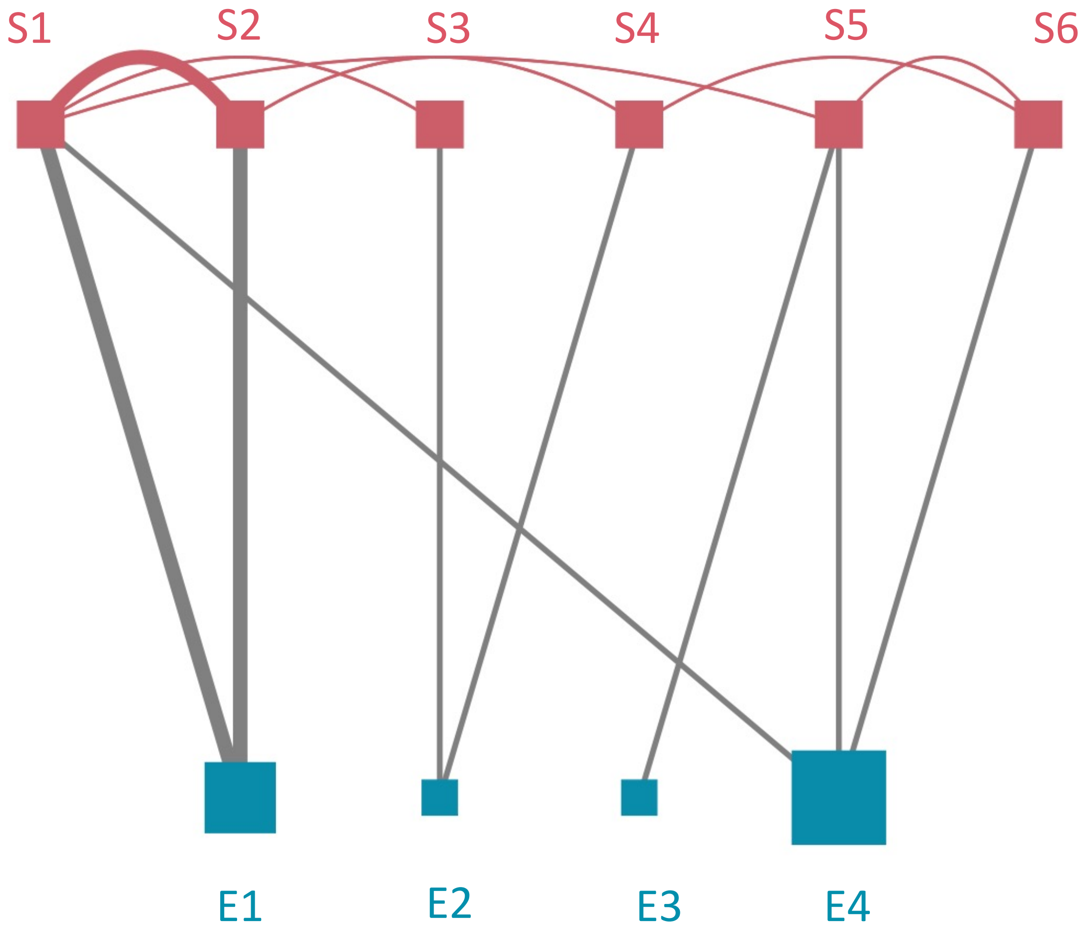
In this small network, there are 3 socio-ecological triangles:

[S1, E1, S2]

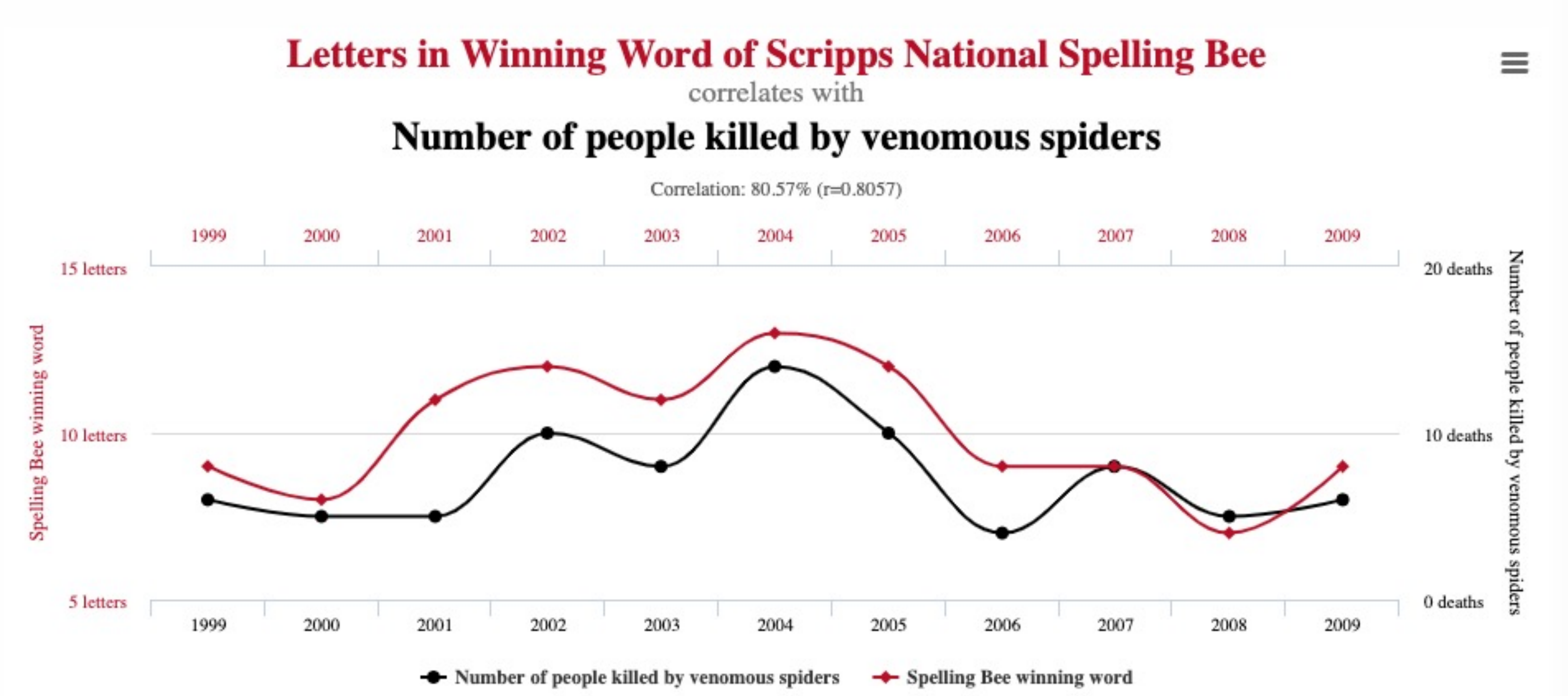
[S1, E4, S5]

[S5, E4, S6]

These are associated with higher outcomes for ecological nodes E1 & E4

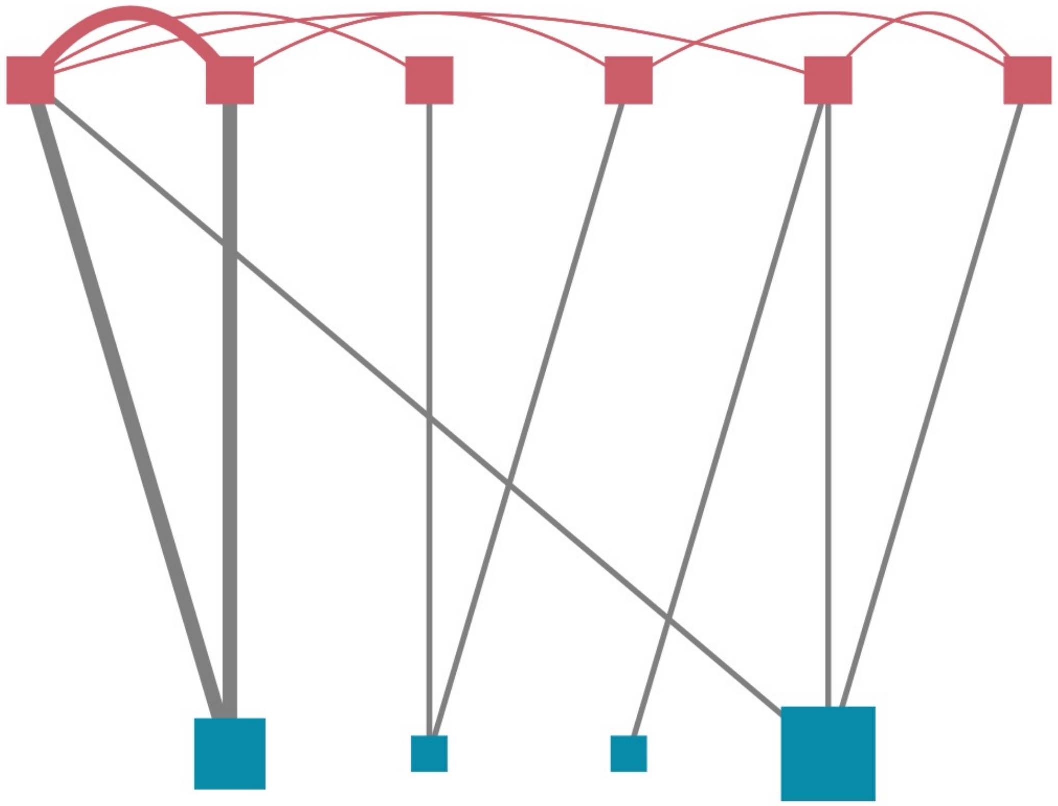


But this co-relationship could be spurious.



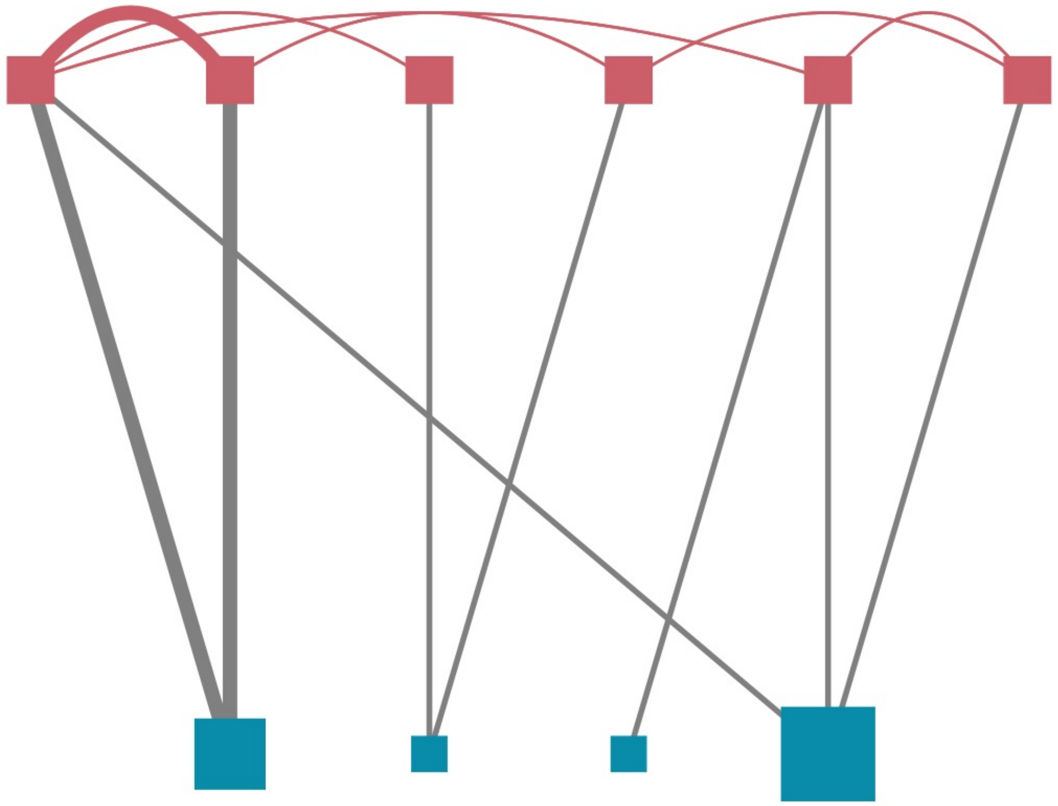
Statistical Social Network Analysis

If the proposed network mechanism were not important, how often would the observed relationship arise by random chance?

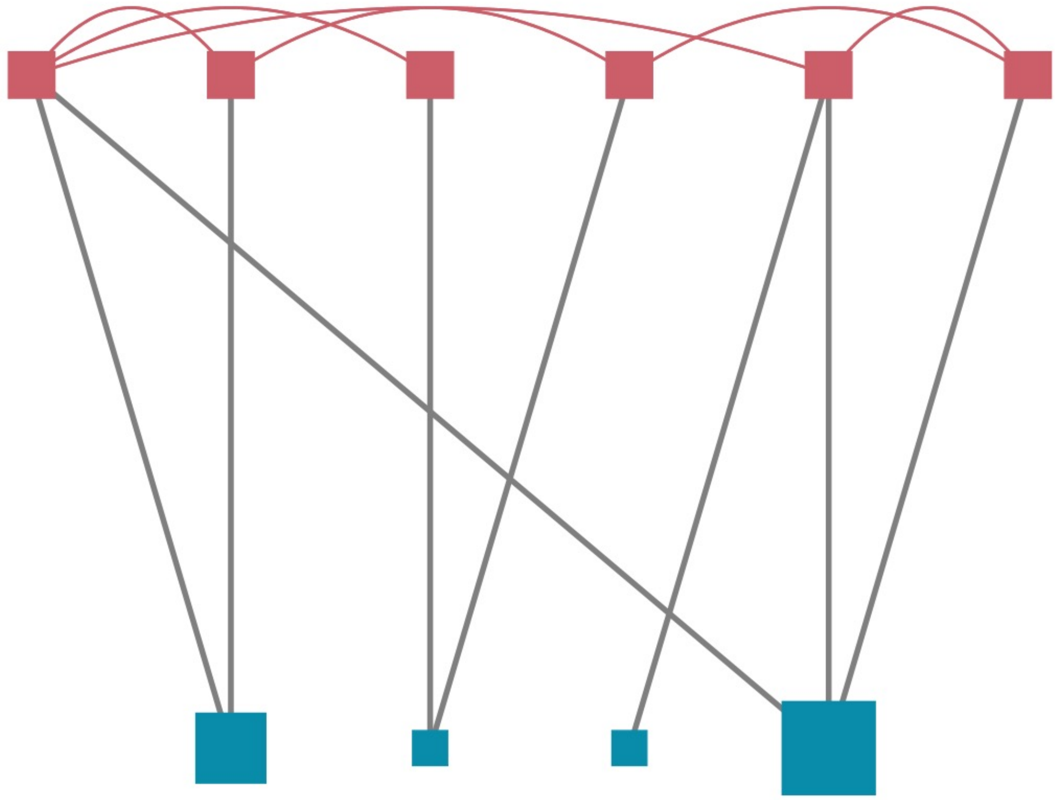


Statistical Social Network Analysis

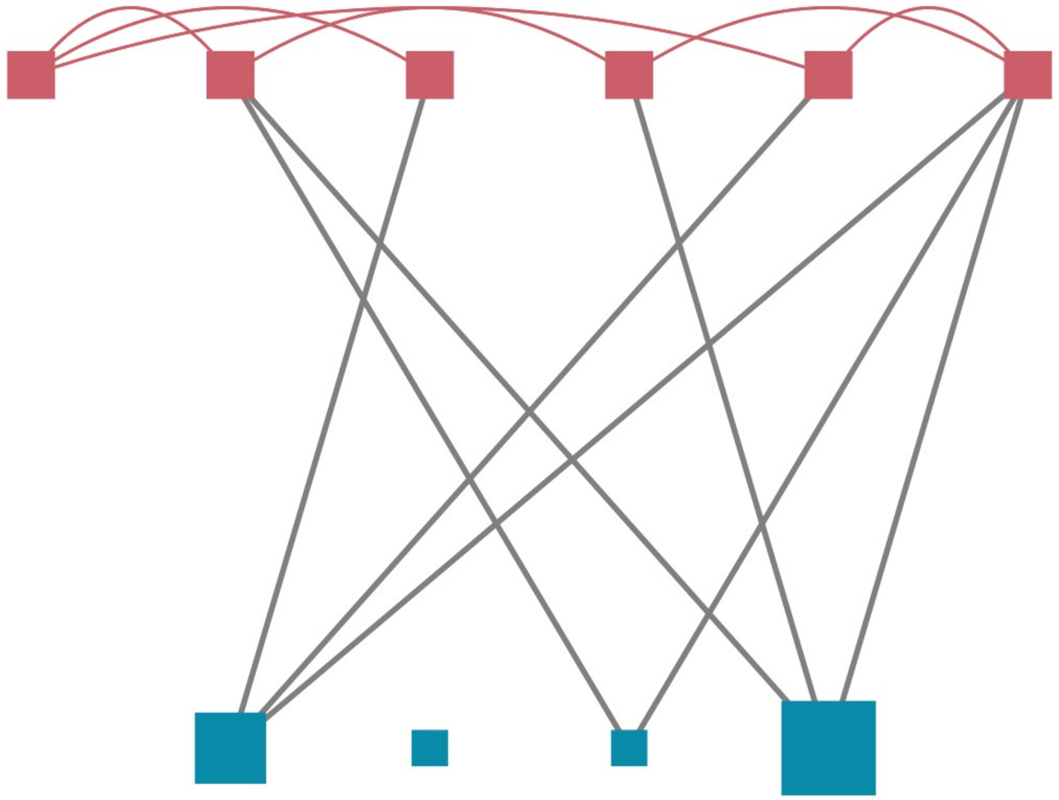
- We need to create a null hypothesis for how often we would observe this relationship at random.
- Let's re-shuffle the connections between the nodes at random, many times.
- This randomization implies that the SES connections are irrelevant to the ecological outcomes



Statistical Social Network Analysis

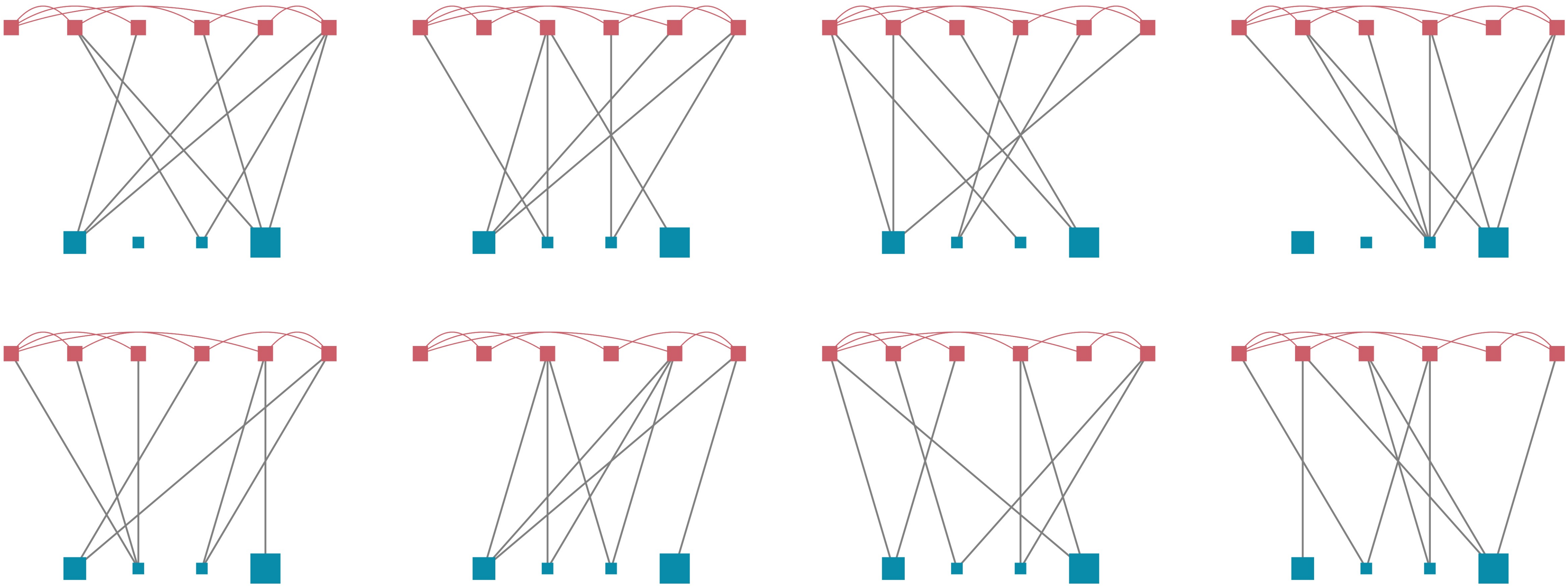


Here is the original network



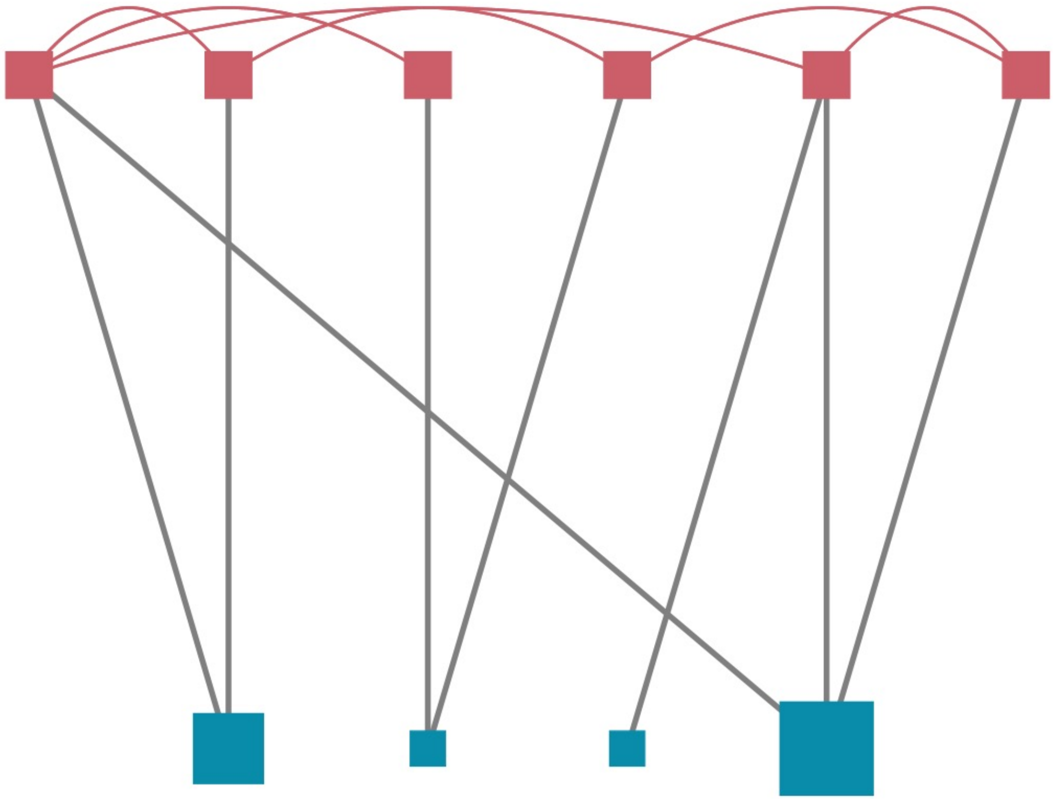
Here is the network with randomised edges

Statistical Social Network Analysis

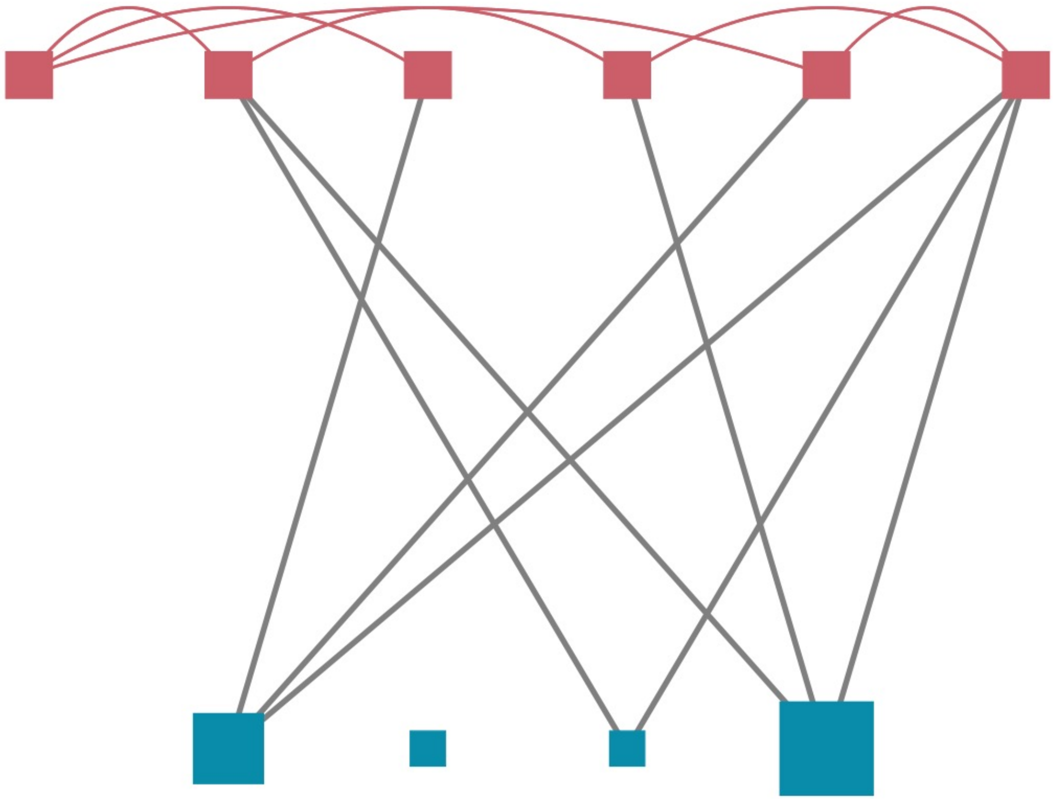


Eight networks with randomised edges

Statistical Social Network Analysis



Here is the original network
There are 3 socioecological triangles

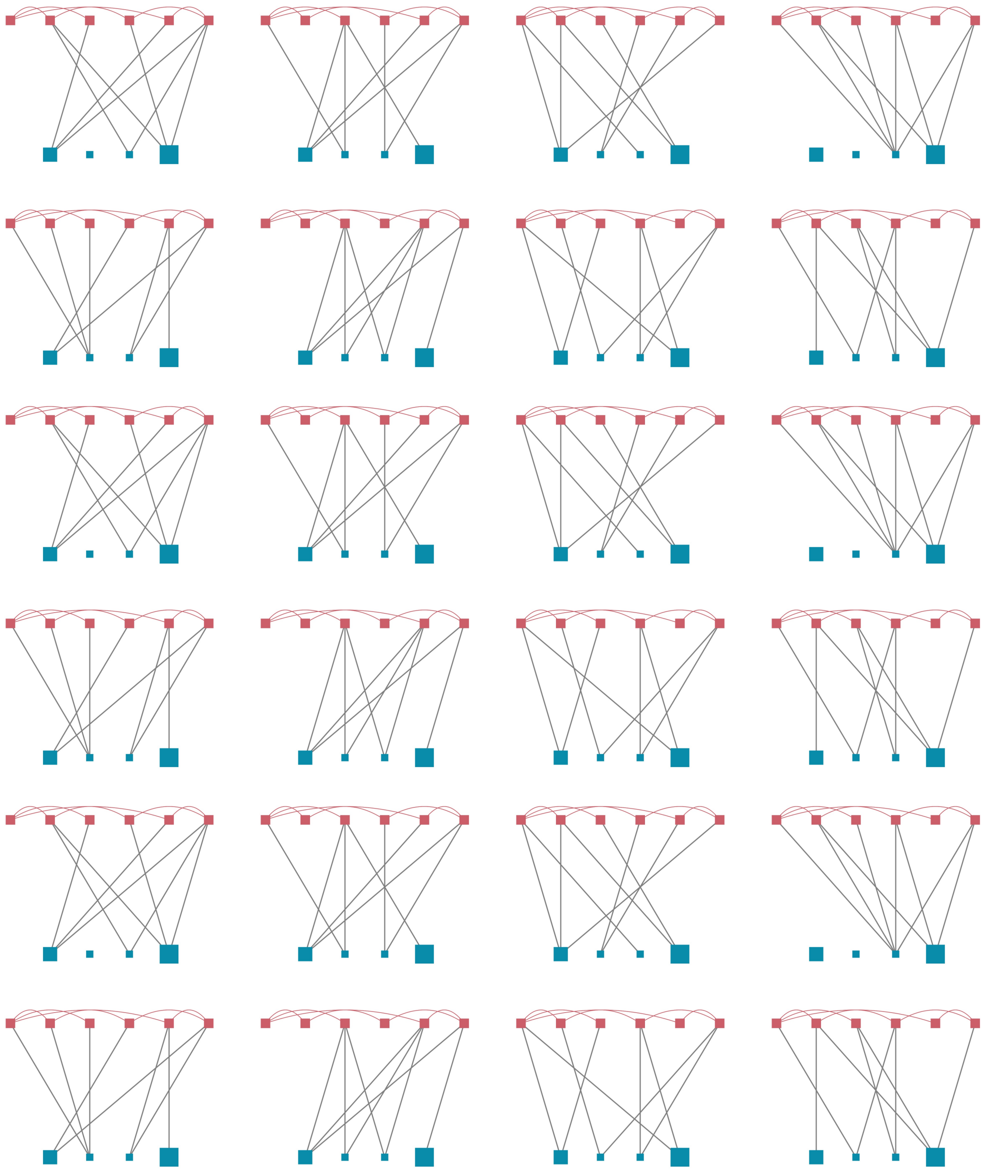


Here is the network with randomised edges
There are also 3 socioecological triangles

Statistical Social Network Analysis

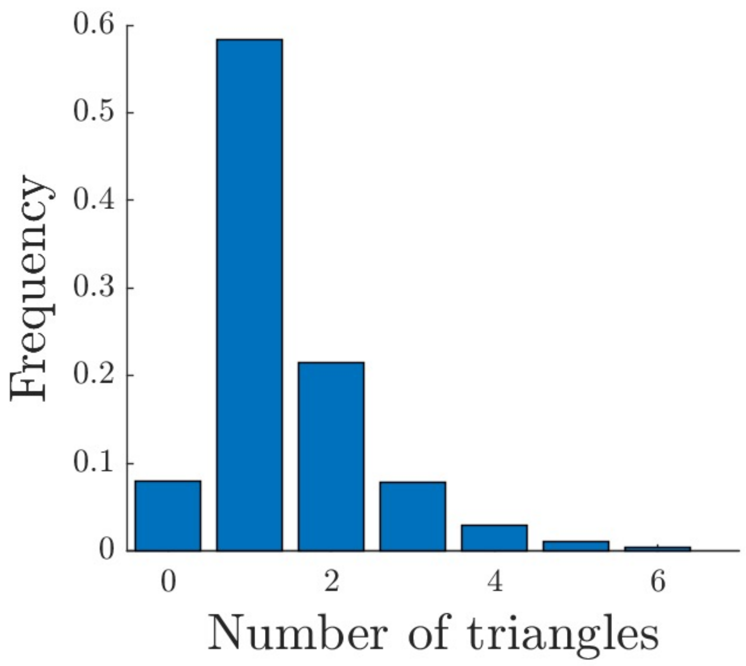
Across thousands of randomised networks:

- How many triangles do I observe at random?
- How many triangles are associated with the two good ecological outcomes, relative to the two poor ecological outcomes?

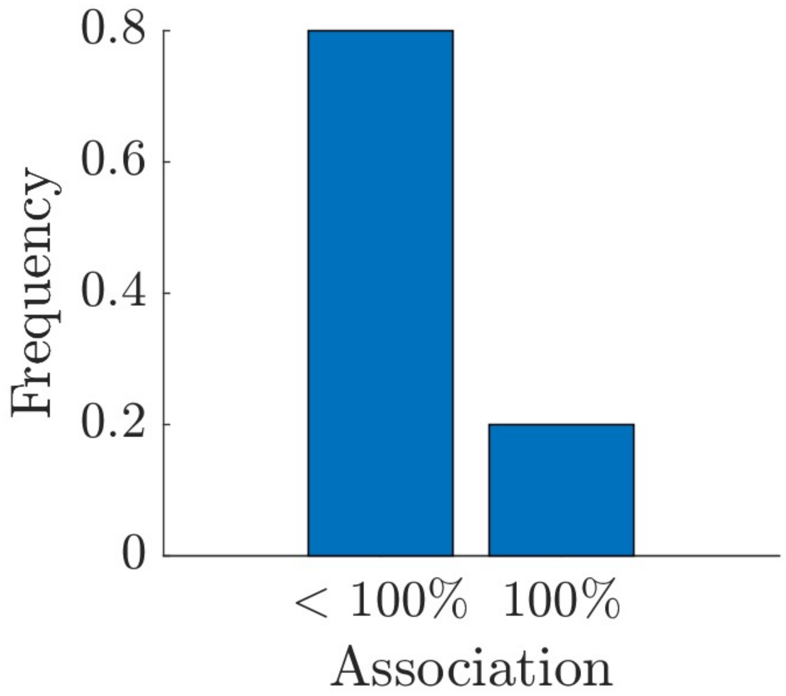


Statistical Social Network Analysis

How many triangles do I observe at random?



How often are triangles 100% associated with positive ecological outcomes?



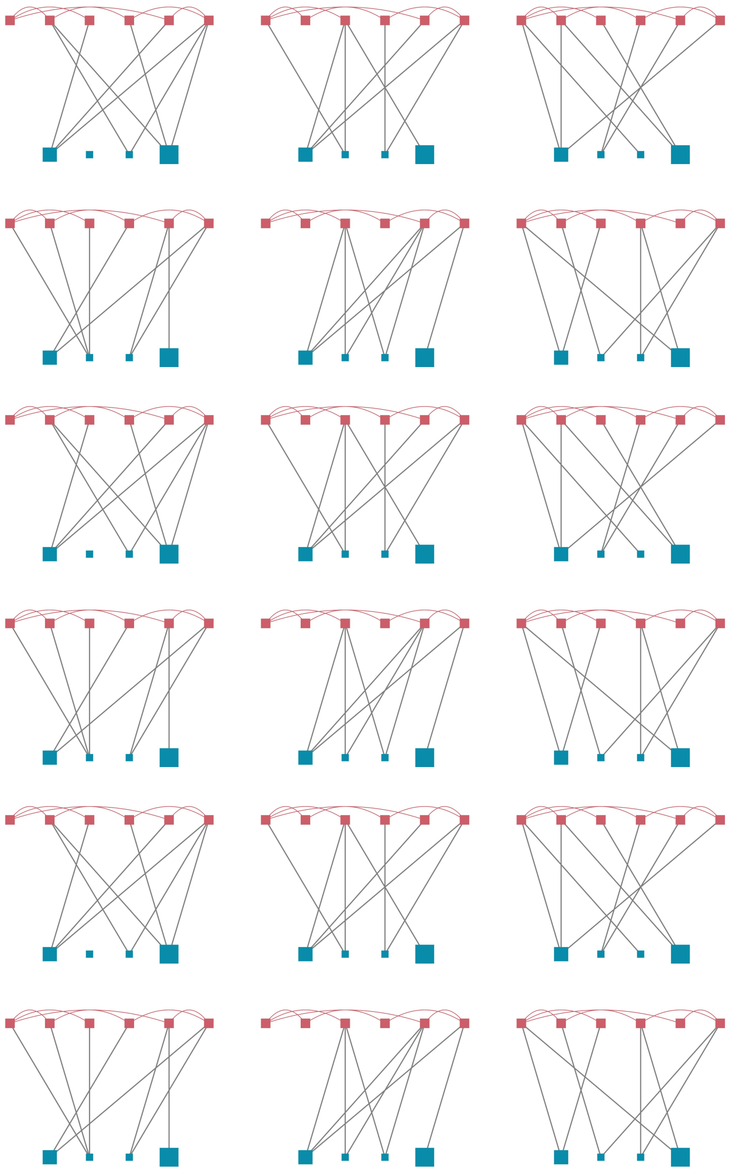
Exponential family random graph models (ERGMs)

ERGMs were developed in response to the type of null model we just used.

They argue that random network generation removes too much structure.

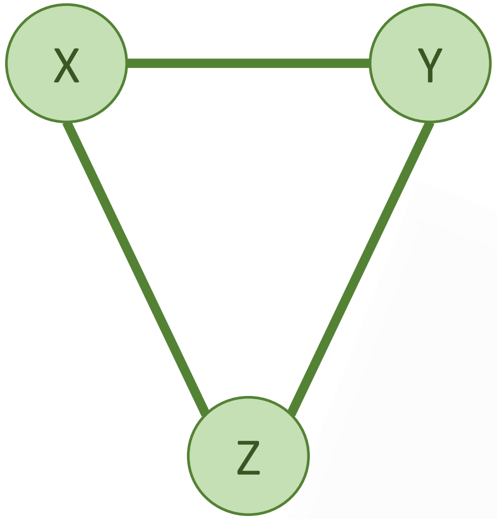
- E.g., Our null should have lots of triangles in it.
- E.g., Our null should have particular in-degree distributions (more high-degree nodes)

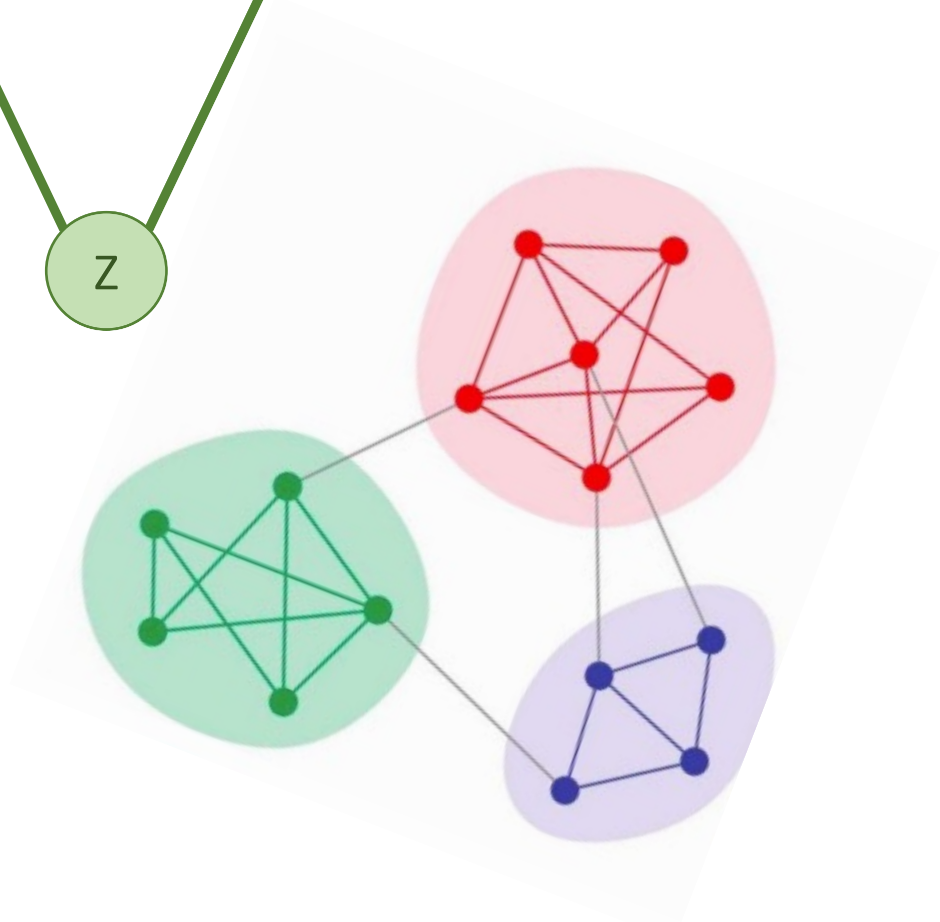
ERGMs attempt to produce many networks that contain these properties.



Modularity in networks & graphs

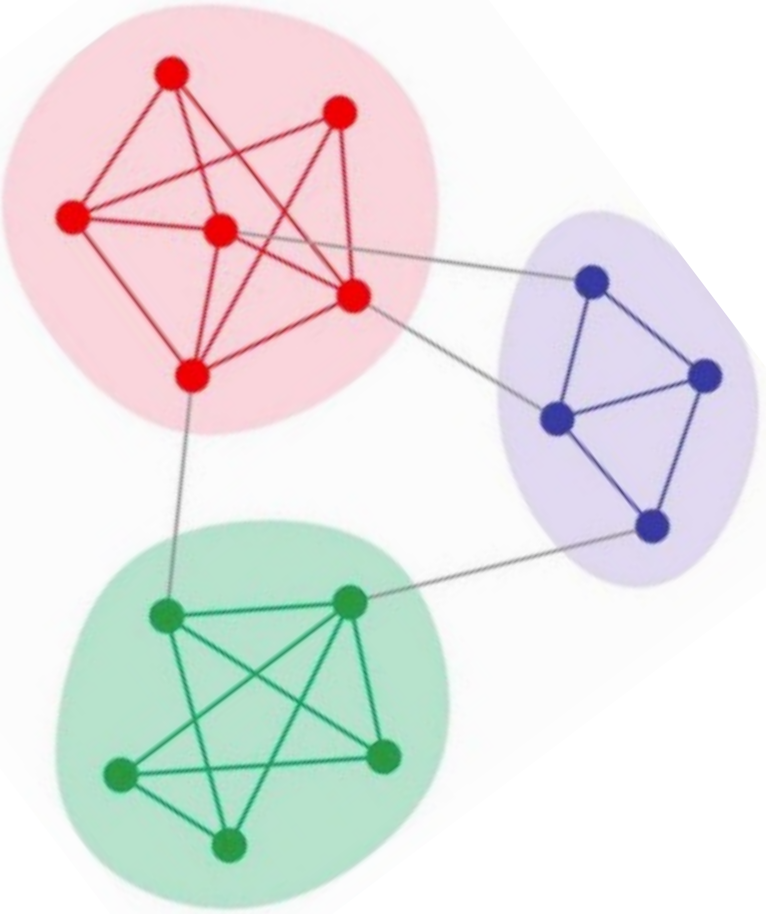
- Network metrics attempt to summarise structure in node-specific or whole-network metrics.
- Sometimes metrics try to capture local neighbourhood properties (e.g., clustering)
- This is challenging when the neighbourhood structure involves large subsets of the network
- “Modularity” is an example of such mesoscale structure.





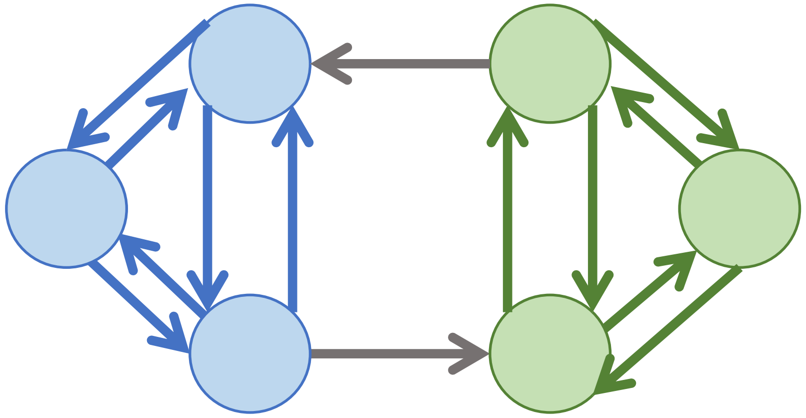
Modularity in networks & graphs

- Modularity is a metric associated with a "partition" of a graph into subgraphs
- Modularity measures the relative strengt of connections within a subgraph, compared to the strength of connection between subgraphs.
- Modularity measures the number of connections within a subgraph, compared to the number you would expect if the edges were allocated at random.



Modularity in networks & graphs

- Imagine we had a network of six nodes with 14 directed edges.
- We propose a partition of the graph into two modules
- If edges were distributed at random, then there would be an average of 2.3 connections entirely within each module.
- There are in fact 6 connections within each. That sounds like two strong modules.



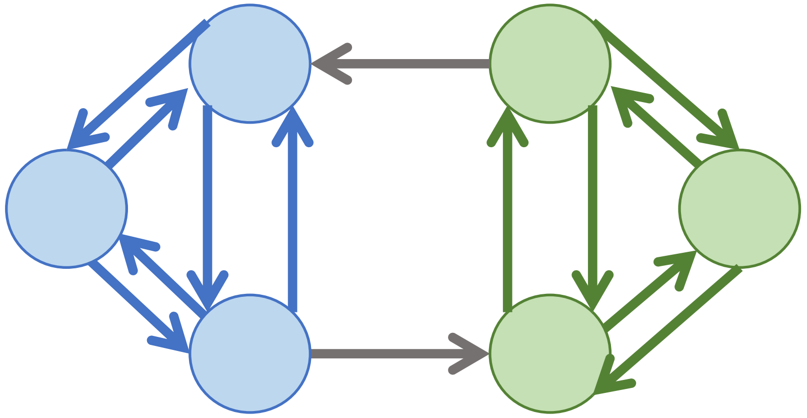
Modularity in networks & graphs

We could propose that:

$$Q = \sum_i \frac{e_{ii}}{n_i x}$$

where:

$$x = \frac{L}{N}$$



e_{ii} is the number of edges entirely within module i

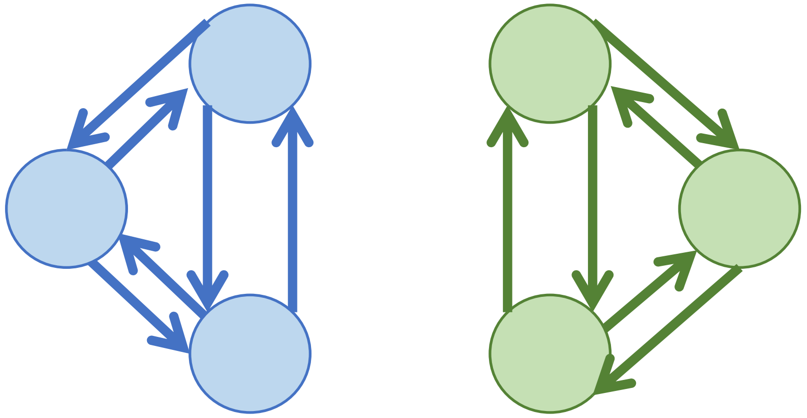
n_i is the number of nodes in module i

L is the total number of edges

N is the total number of nodes

Modularity in networks & graphs

What's the modularity of this network?



$$Q = \sum_i \frac{e_{ii}}{n_i x}$$

where:

$$x = \frac{L}{N}$$

e_{ii} is the number of edges entirely within module i

n_i is the number of nodes in module i

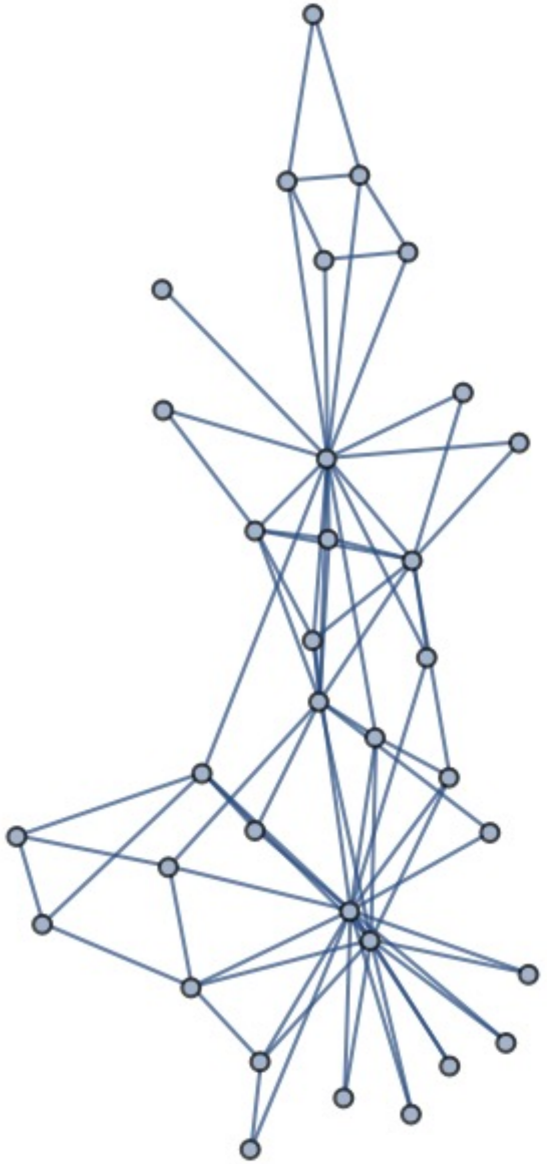
L is the total number of edges

N is the total number of nodes

Identifying modular partitions

Let's look at the Zachary Karate club dataset

- Social network of 34 Karate club members 1970-72
- Group split into two groups following friction between two key members
- Communication network had been mapped beforehand
- Division membership was accurately predicted by network structure.



Zachary (1977) "An information flow model for conflict and fission in small groups." *Journal of Anthropological Research*.

Modularity in networks & graphs

- There are ***lots*** of different ways to partition a graph into modules
- For a 5 node graph, there are 52 potential partitions
- For a 12 node graph, there are 4,213,597 potential partitions
- It would be time-consuming to calculate the modularity of each of these partitions, and then choose the maximum.