Graphs, Food Webs and Biodiversity

Midge Cozzens, Rutgers University
MBI
July 29, 2013

Indicators of Biodiversity

- One of the key goals of conservation biologists is preservation of biodiversity.
- But what is biodiversity?
- Long history of trying to define it.
- It is a multidimensional concept.

Indicators of Biodiversity

- Traditional approaches to biodiversity consider:
  - *Richness* = number of species
  - *Evenness* = extent to which species are equally distributed
  - *Indicator species* = organism that serves as a measure of environmental conditions

Competition Graphs

- We will try to derive the "dimensions" of biodiversity starting from properties of ecosystems, in particular normal, healthy competition between species.
- Using food webs and *competition graphs*
  - Arose from a problem of ecology.
  - Joel Cohen 1968
- Key idea: Two species compete if they have a common prey.
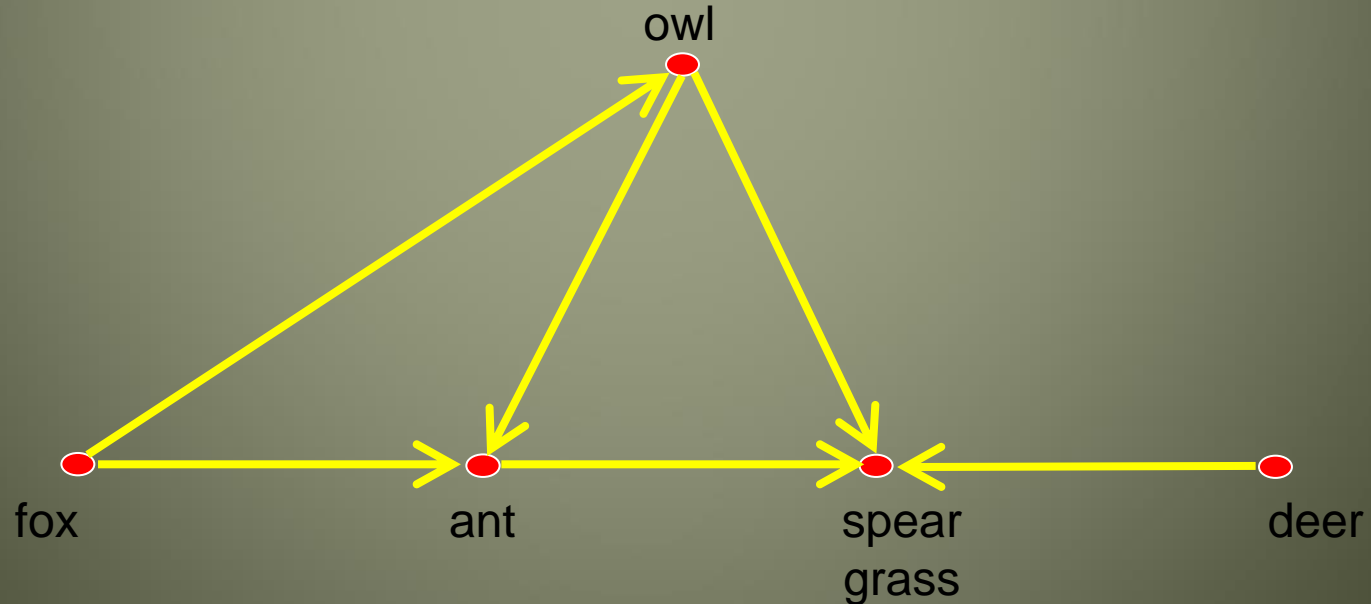
Competition Graphs of Food Webs

Food Webs

      Let the vertices of a directed graph (digraph) be species in an ecosystem.

      Include an arc from x to y if x preys on y.
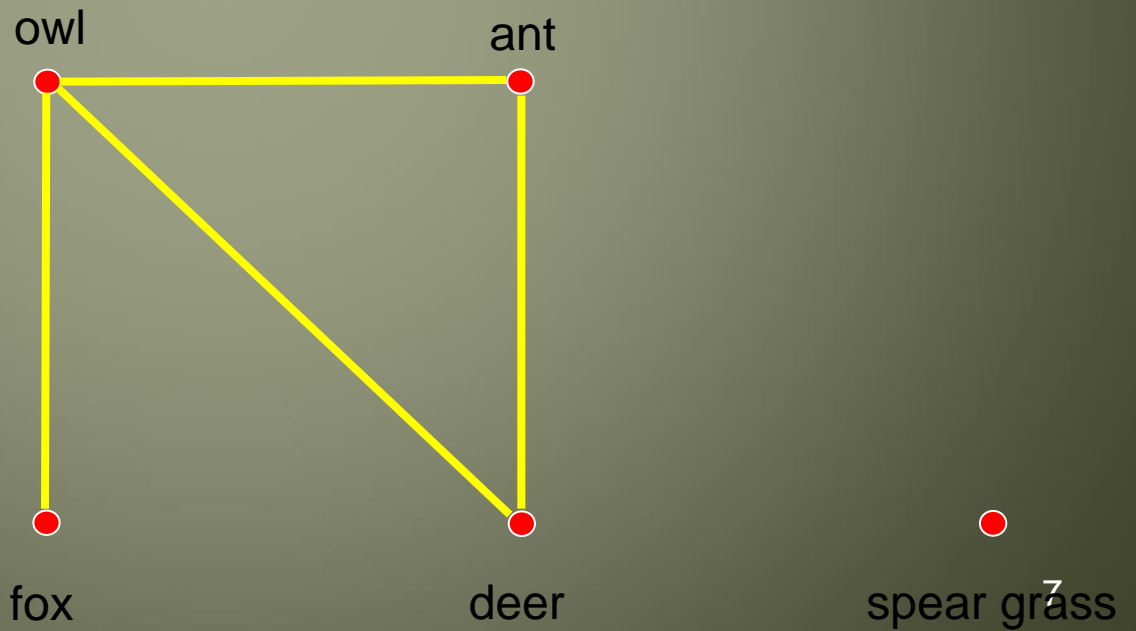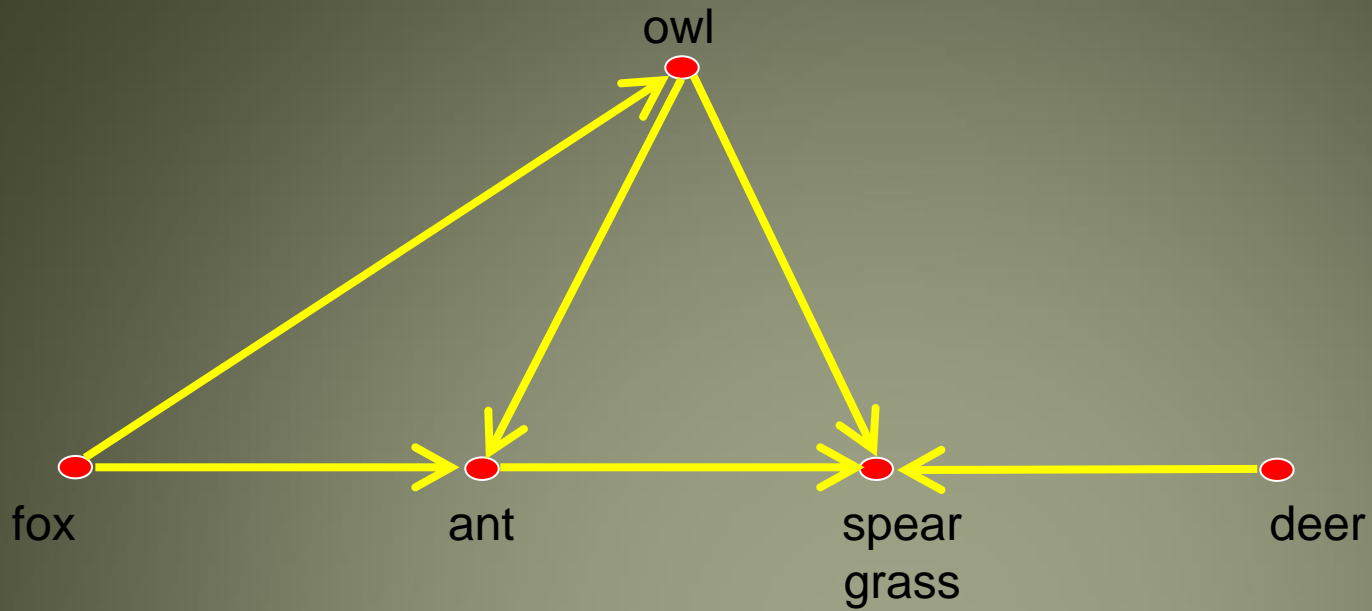
      Usual assumption for us: no cycles.

owl

fox   ant   spear
grass   deer

Competition Graphs of Food Webs

Consider a corresponding undirected graph.

Vertices = the species in the ecosystem

Edge between a and b if they have a common prey, i.e., if there is some x so that there are arcs from a to x and b to x.

owl

fox   ant   spear   deer
                                               grass

owl   ant

fox   deer   spear grass

Competition Graphs

More generally:

Given a digraph D = (V,A)
(Usually assumed to be *acyclic*.)

The ***competition graph*** C(D) has vertex set V and an edge between a and b if there is an x with arcs (a,x) ε A and (b,x) ε A.

Interval Graphs

•A key idea in the study of competition graphs is the notion of interval graph. It arose from a problem in genetics posed by Seymour Benzer.

•*Benzer's Problem* (1959): The problem was: How can you understand the "fine structure" inside the gene without being able to see inside?

Benzer's Problem

•Classically, geneticists had treated the chromosome as a linear arrangement of genes.

•Benzer asked in 1959: Was the same thing true for the "fine structure" inside the gene?

Benzer's Problem

At the time, we could not observe the fine structure directly.

•Benzer studied mutations.

•He assumed mutations involved "connected
•substructures" of the gene.

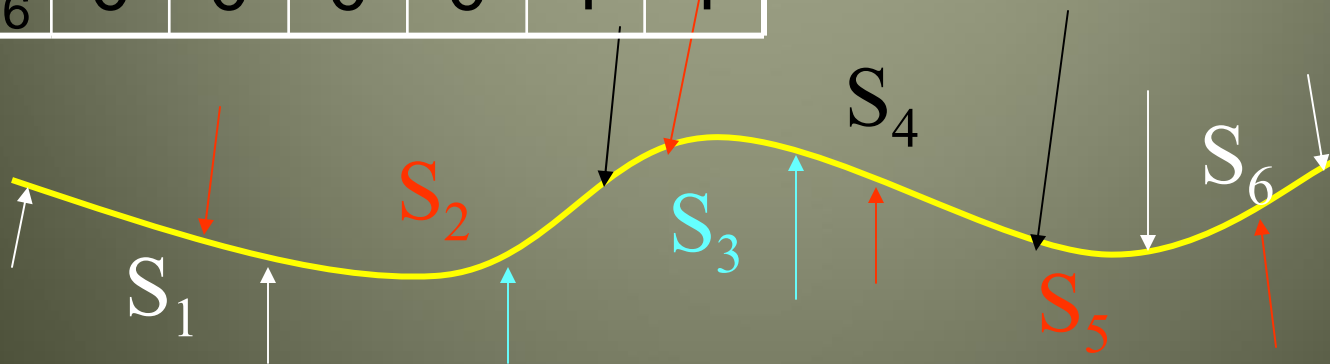By gathering mutation data, he was able to surmise whether or not two mutations overlapped.

Benzer's Problem

	S_1	S_2	S_3	S_4	S_5	S_6
S_1	1	1	0	0	0	0
S_2	1	1	1	1	0	0
S_3	0	1	1	1	0	0
S_4	0	1	1	1	1	0
S_5	0	0	0	1	1	1
S_6	0	0	0	0	1	1

i,j entry is 1 if mutations S_i and S_j overlap, 0 otherwise.

Benzer's Problem

	S_1	S_2	S_3	S_4	S_5	S_6
S_1	1	1	0	0	0	0
S_2	1	1	1	1	0	0
S_3	0	1	1	1	0	0
S_4	0	1	1	1	1	0
S_5	0	0	0	1	1	1
S_6	0	0	0	0	1	1

S_4

S_2

S_3

S_1

S_5

S_6

Benzer's Problem

Can the fine structure of a gene be represented as an interval graph? We say that the graph is an *interval graph* if it is consistent with a linear arrangement.

That is: A graph is an interval graph if we can find intervals on the line so that two vertices are joined by an edge if and only if their corresponding intervals overlap.

• Interval graphs have been very important in genetics. Long after Benzer's problem was solved using other methods, interval graphs played a crucial role in *physical mapping of DNA* and more generally in the *mapping of the human genome.*

14

Benzer's Problem

- Given a graph, is it an interval graph?
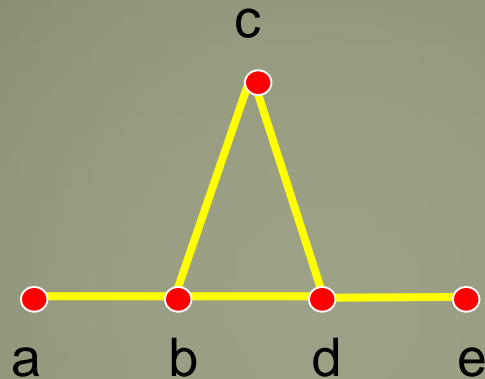
- We need to find intervals on the line that have the same overlap properties
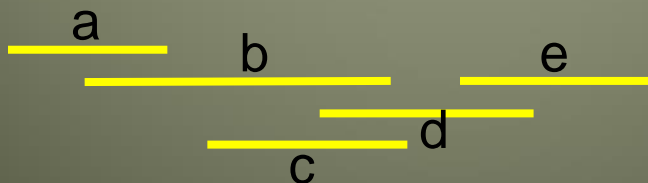
Benzer's Problem

• The following is not an interval graph.
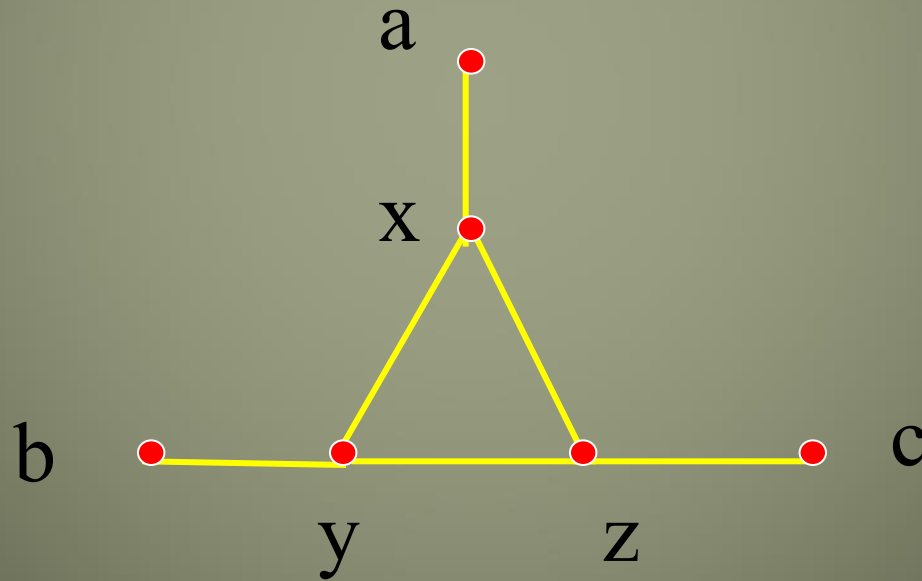
Benzer's Problem

The following is not an interval graph. Once we give intervals for a, b, c, y and z there is no room for x without overlapping b.
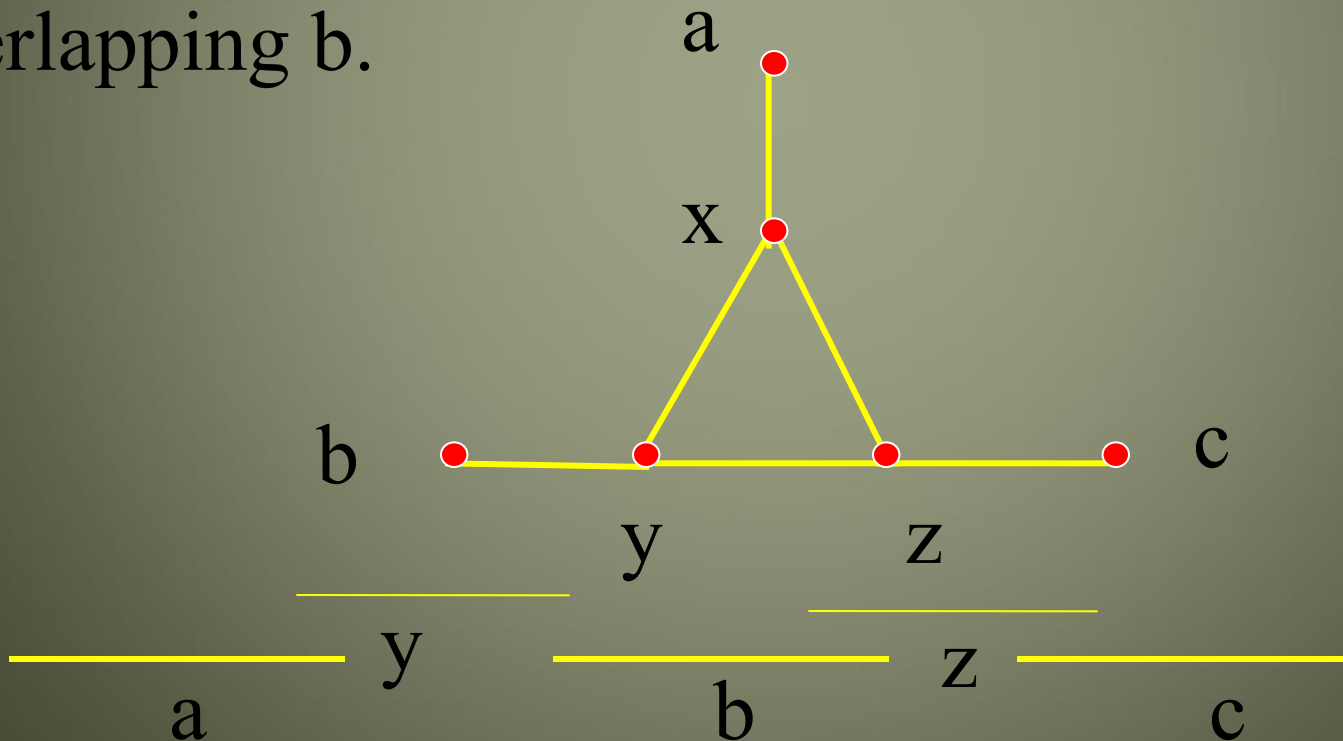
Intersection of Boxes

More generally, we can study ways to represent graphs where the edges correspond to intersections of boxes in Euclidean space.

The *boxicity* of G is the smallest p so that we can assign to each vertex of G a box in Euclidean p-space so that two vertices are neighbors iff their boxes overlap.
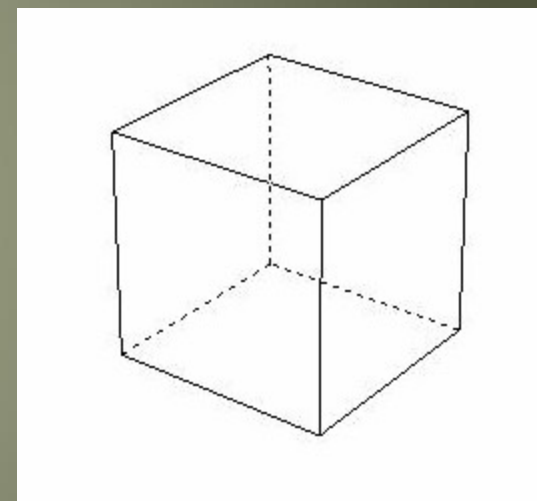
Well-defined (Roberts 1968) but hard to compute (Cozzens 1980).

Intersection of Boxes

- Interval graphs are the graphs of boxicity 1.

- Consider the graph C_4.

- It is not an interval graph.

- However, it can be represented as the intersection graph of boxes in 2-space.

- So, boxicity of C_4 is 2.

$G = C_4$

a   b

d   c
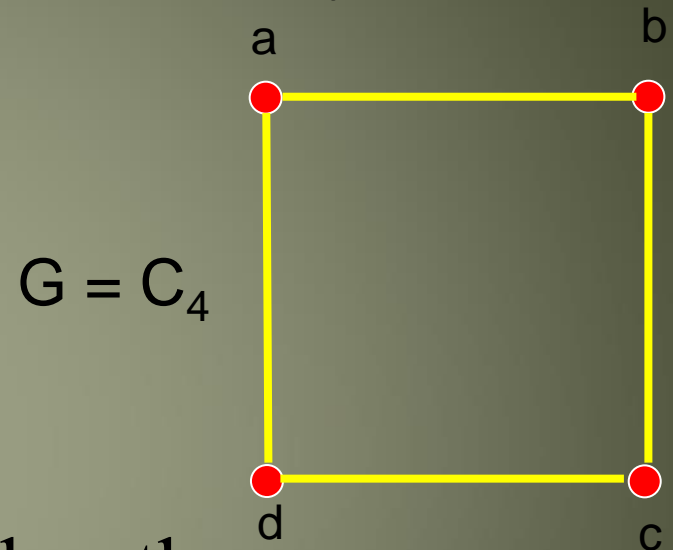
Intersection of Boxes

•C_4 can be represented as the intersection graph of boxes in 2-space.

•So, boxicity of C_4 is 2.

$G = C_4$

a   b

d   c

b

a   c

d
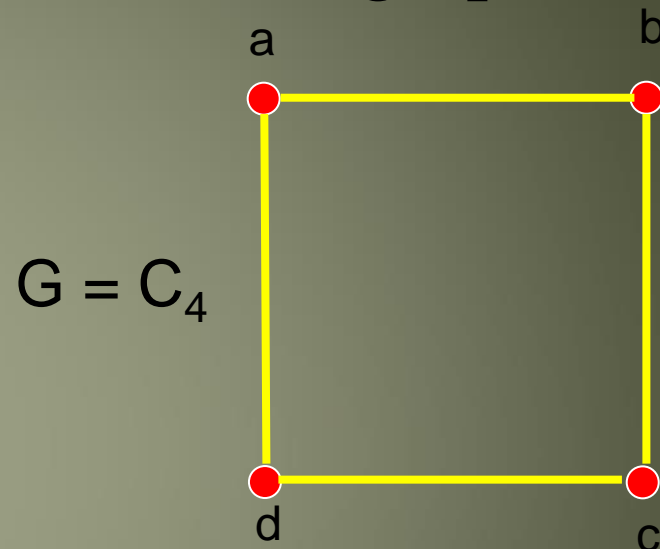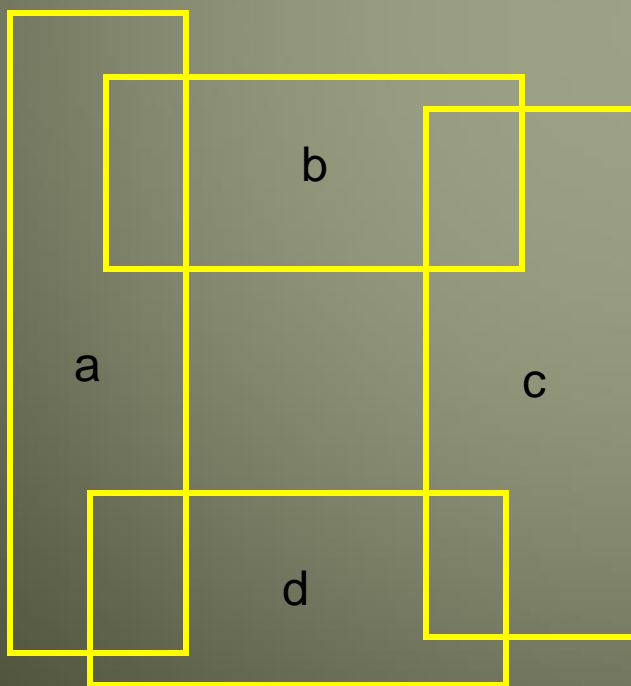
Factors Affecting Biodiversity

- Different factors determine a species' normal healthy environment.
  - Moisture
  - Temperature
  - pH
  - …

- We can use each such factor as a dimension.
- Then the range of acceptable values on each dimension is an interval.
- Each species can be represented as a box in Euclidean space.
- The box represents its ecological niche.

Factors Affecting Biodiversity
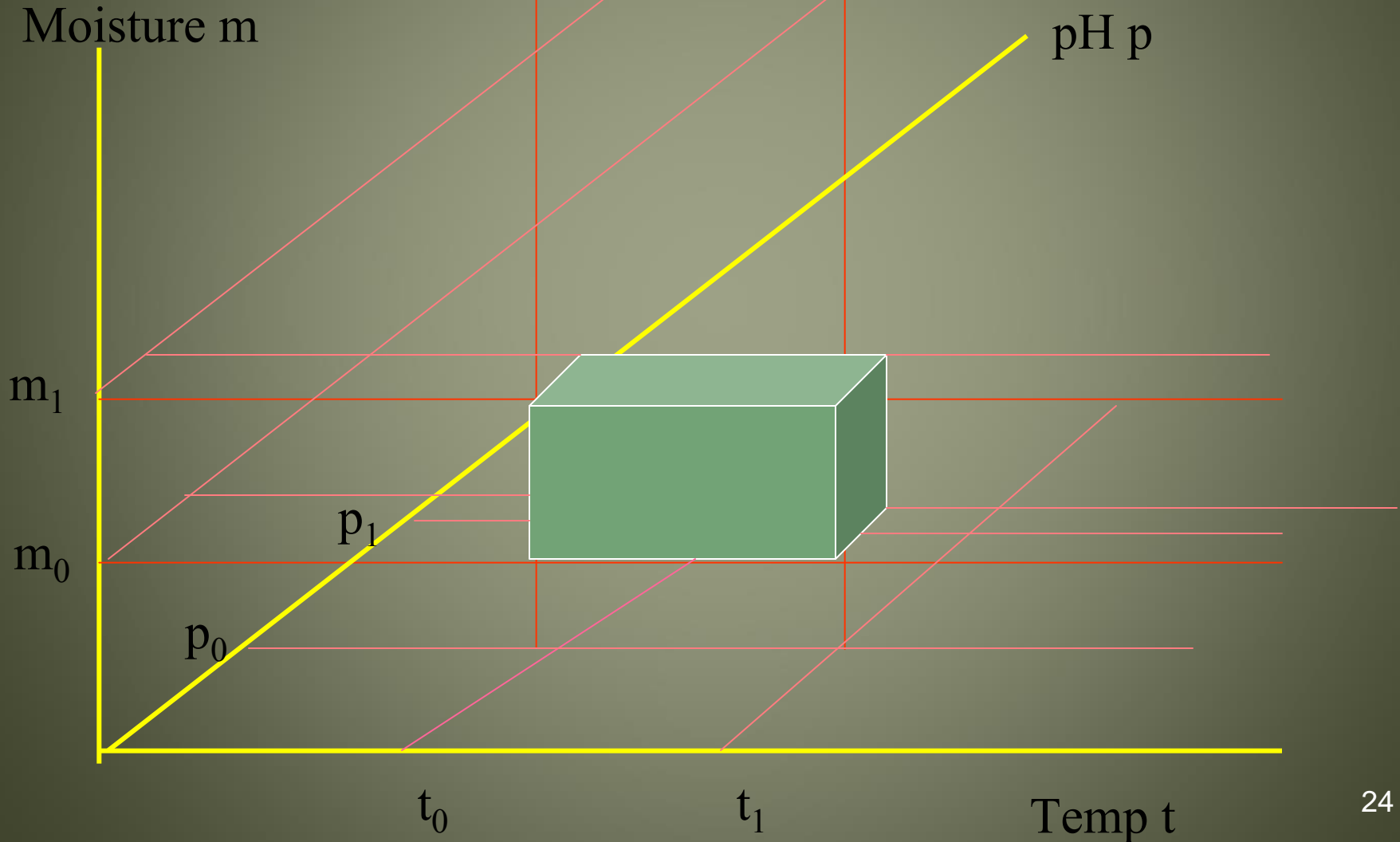
- The ecological niche is a box.

Moisture m

m_1

m_0

t_0 t_1 Temp t
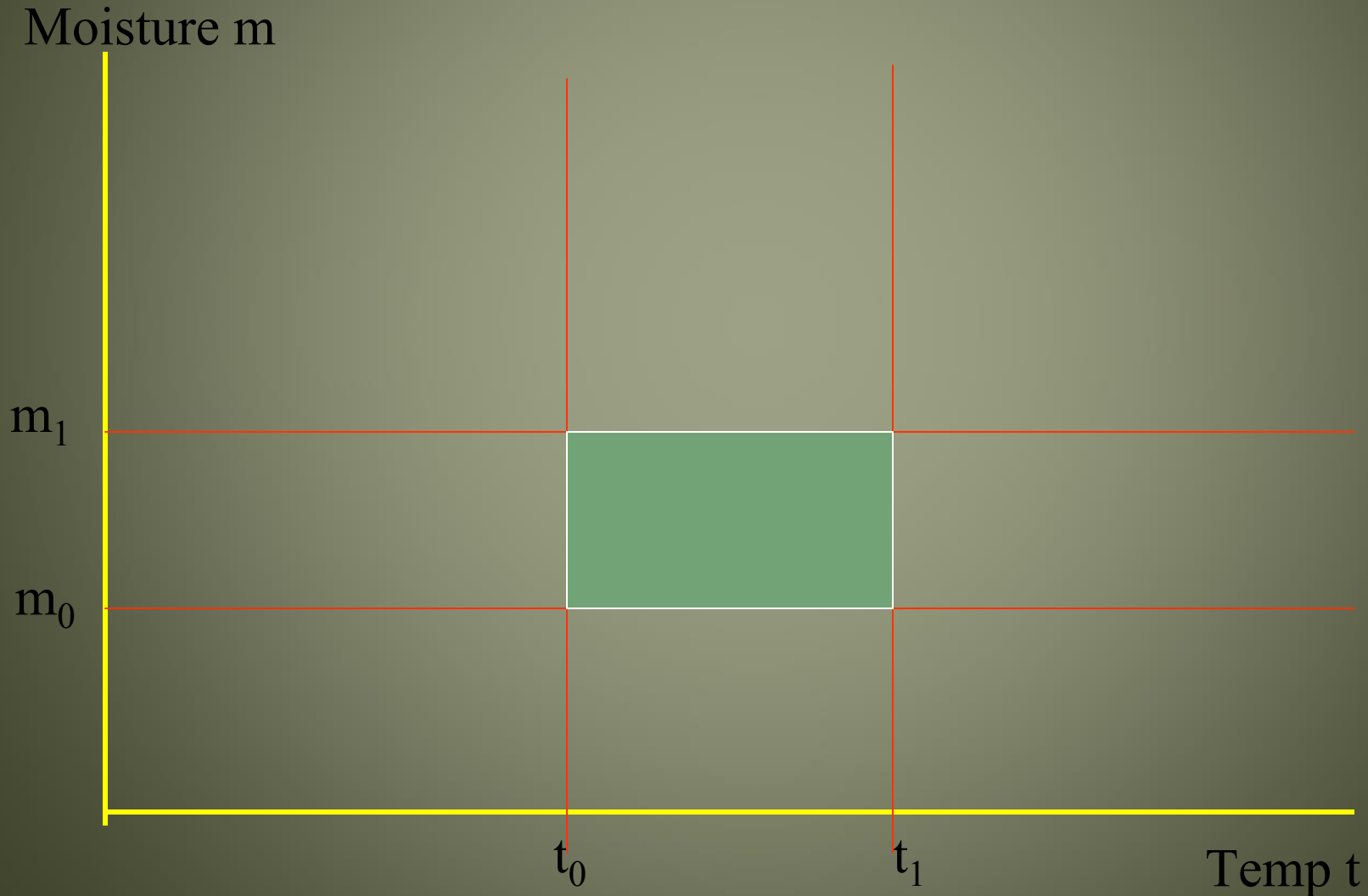
Factors Affecting Biodiversity

• Simplifying assumption: acceptable ranges on each dimension are independent of values on other dimensions.

Moisture m

m_1

m_0

t_0 t_1 Temp t

23

Factors Affecting Biodiversity

•The ecological niche is a box.

Moisture m

pH p

m_1

m_0

p_1

p_0

t_0　　　　t_1　　　　Temp t
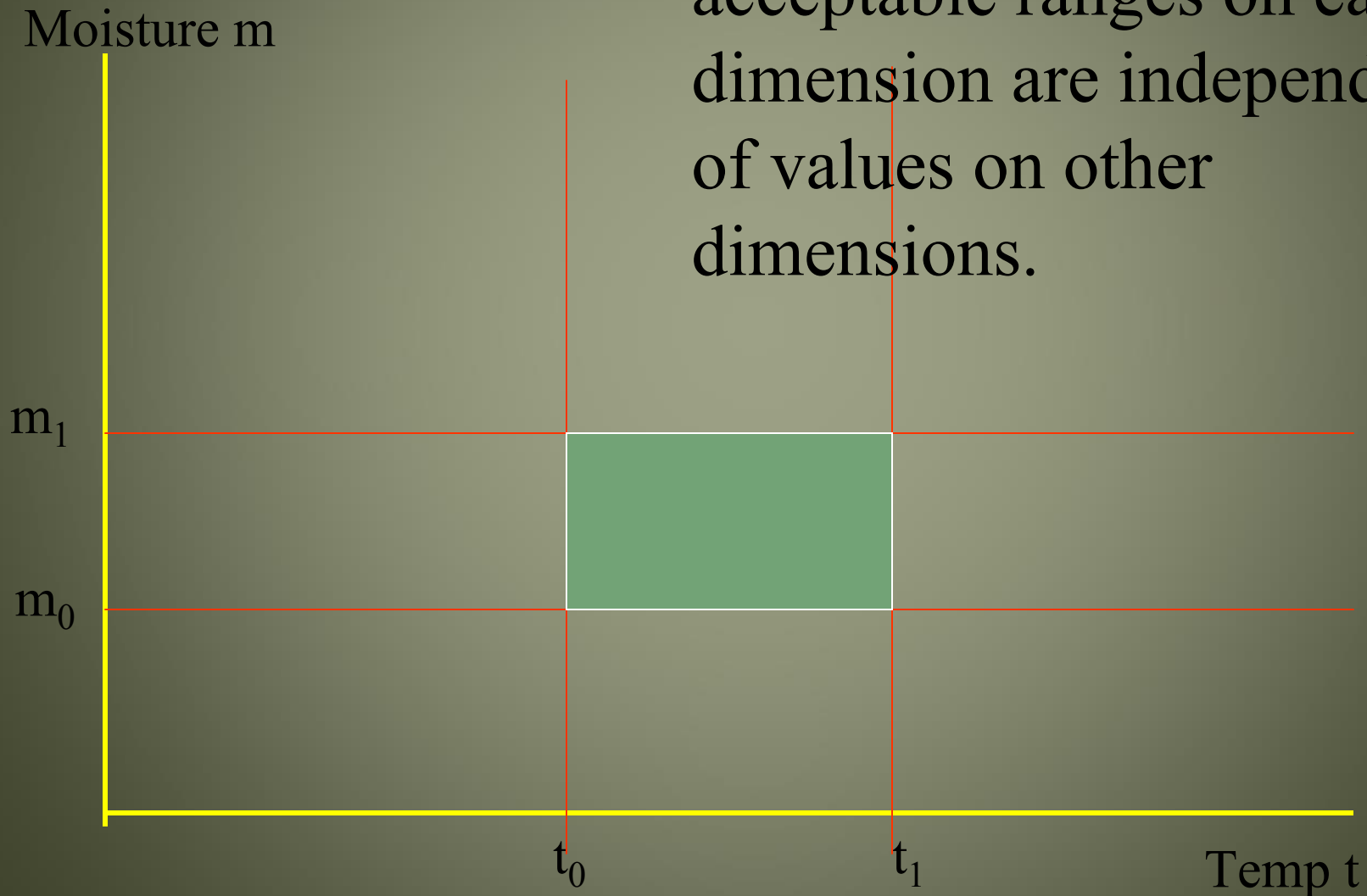
Competition

- Old ecological principle: Two species compete if and only if their ecological niches overlap.

- Joel Cohen (1968):
    - Start with an independent definition of competition
    - Map each species into a box (niche) in k-space so competition corresponds to box overlap (niche overlap)
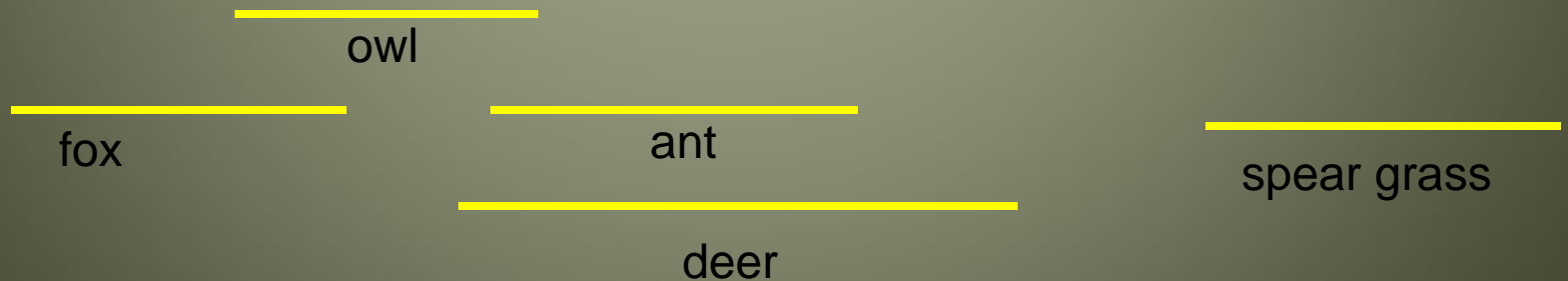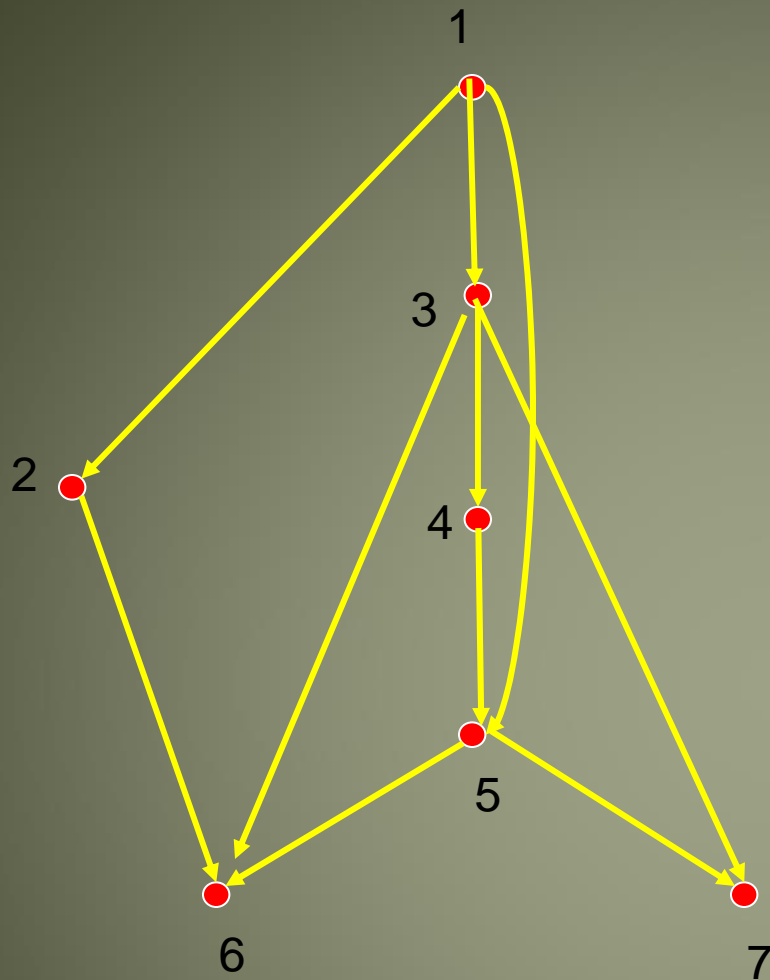   - Find smallest k that works.

Competition

•Specifically, Cohen started with the competition graph as defined before.

•The question then becomes: What is the boxicity of the competition graph?

owl

fox   ant   spear   deer
                                              grass

owl   ant

fox   deer   spear grass
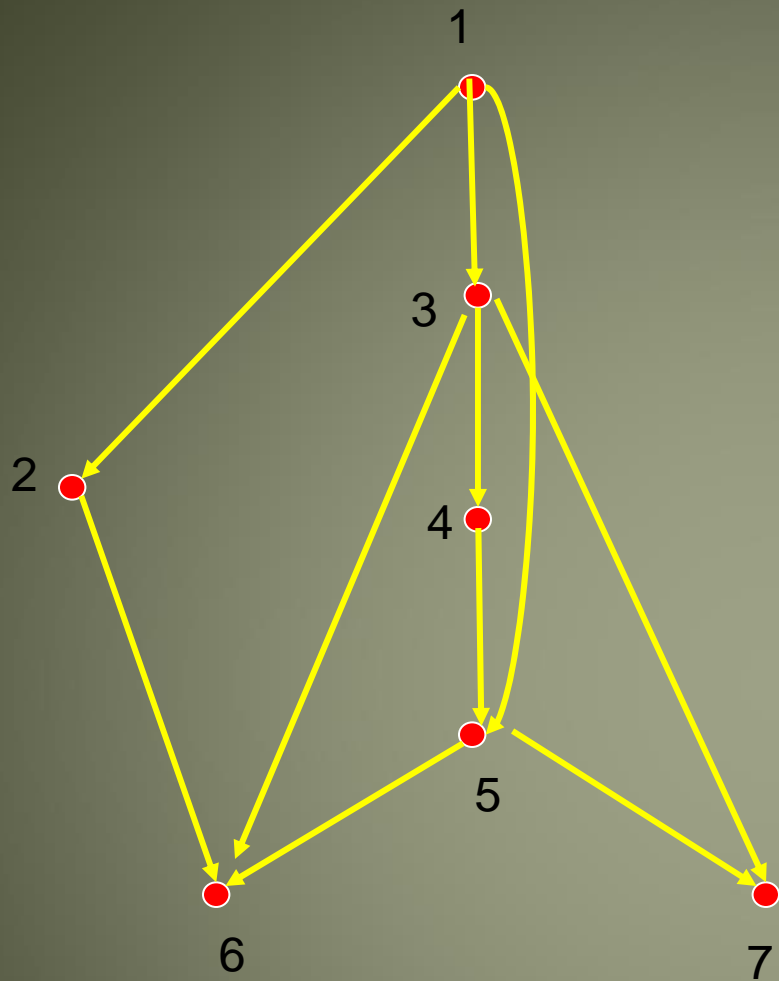
27
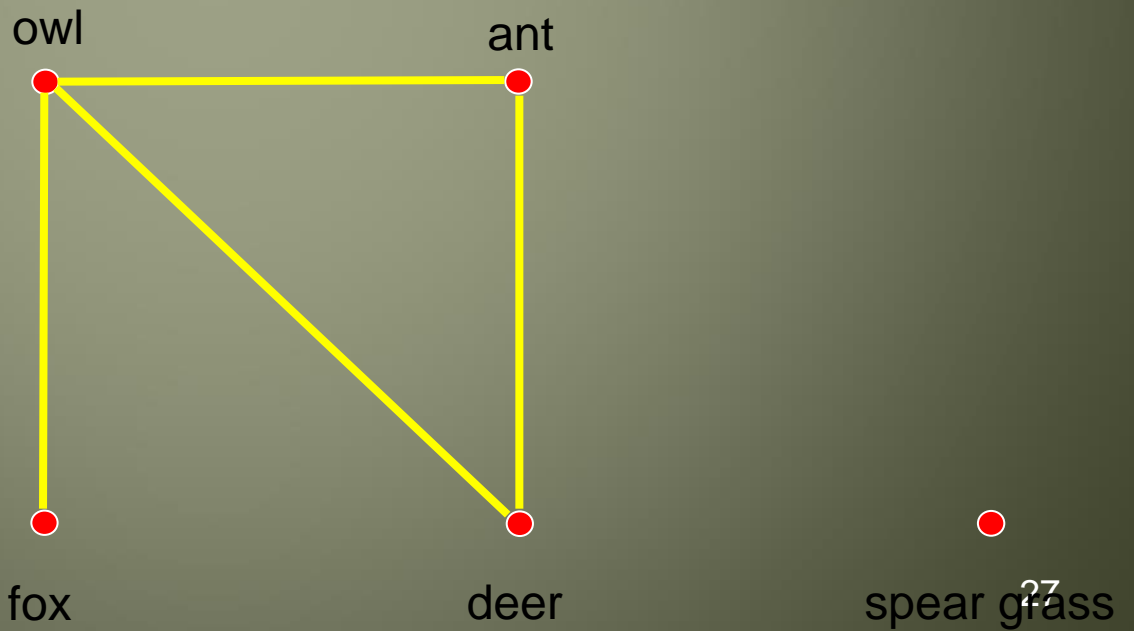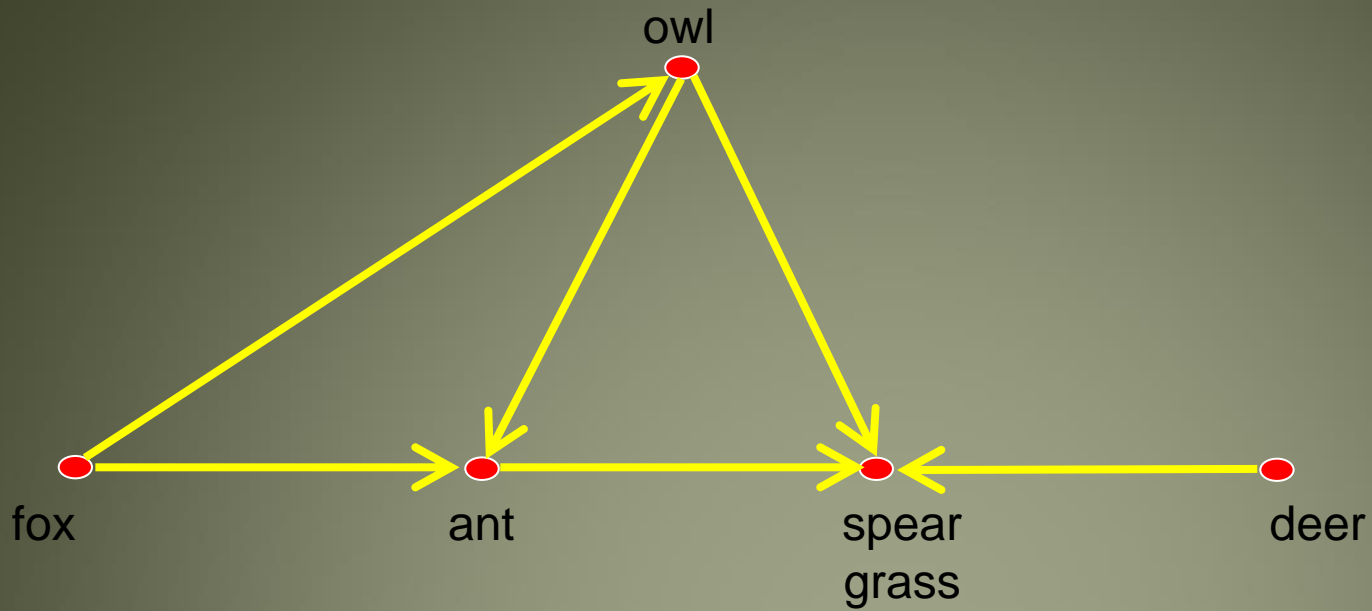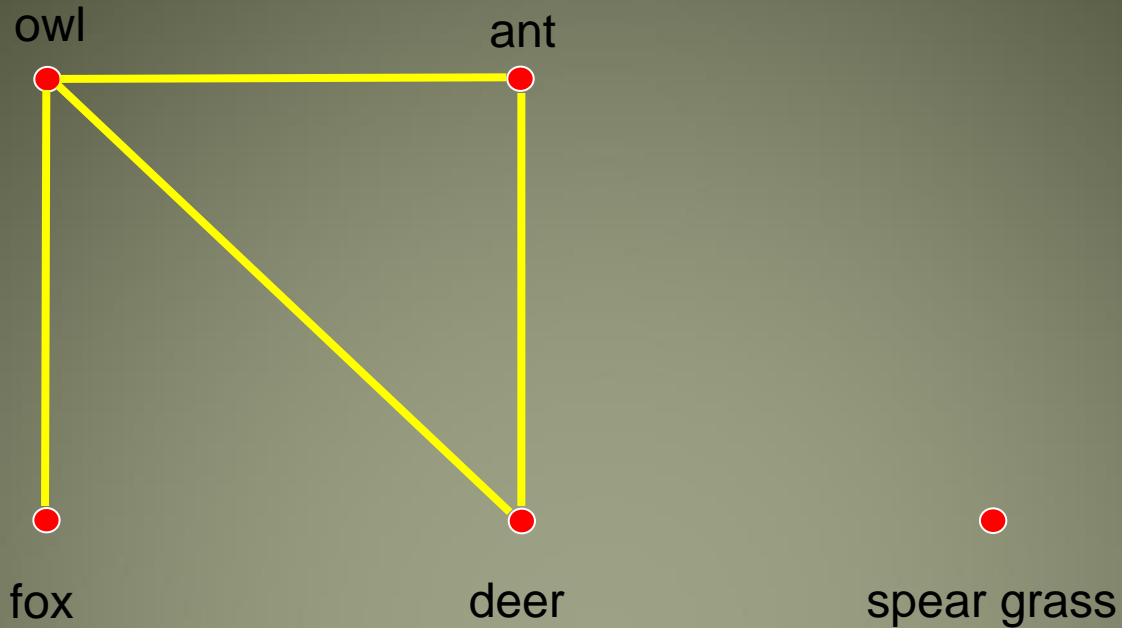
This is an interval graph. Thus, boxicity is 1.

Key:
1. Juvenile Pink Salmon
2. P. Minutus
3. Calanus & Euphasiid Barcillia
4. Euphasiid Eggs
5. Euphasiids
6. Chaetoceros Socialis & Debilis
7. Mu-Flagellates
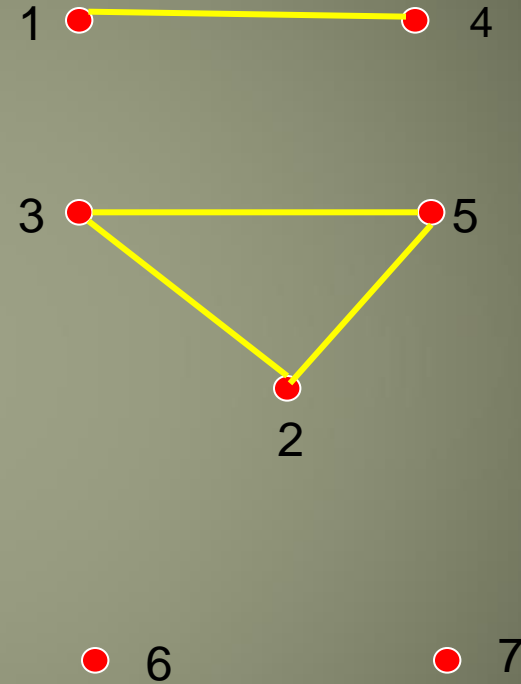
Strait of Georgia, British Columbia, Canada
Due to Parsons and LeBrasseur
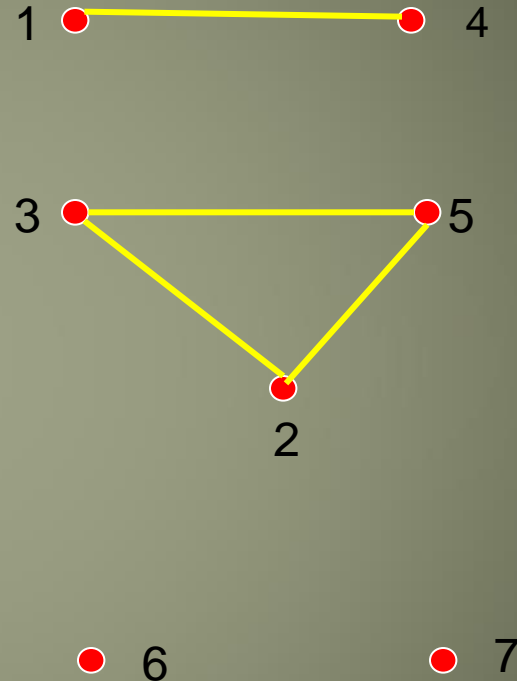From Joel Cohen, *Food Webs and Niche Space*
Princeton University Press, 1978

29

Competition graph

Strait of Georgia, British Columbia, Canada

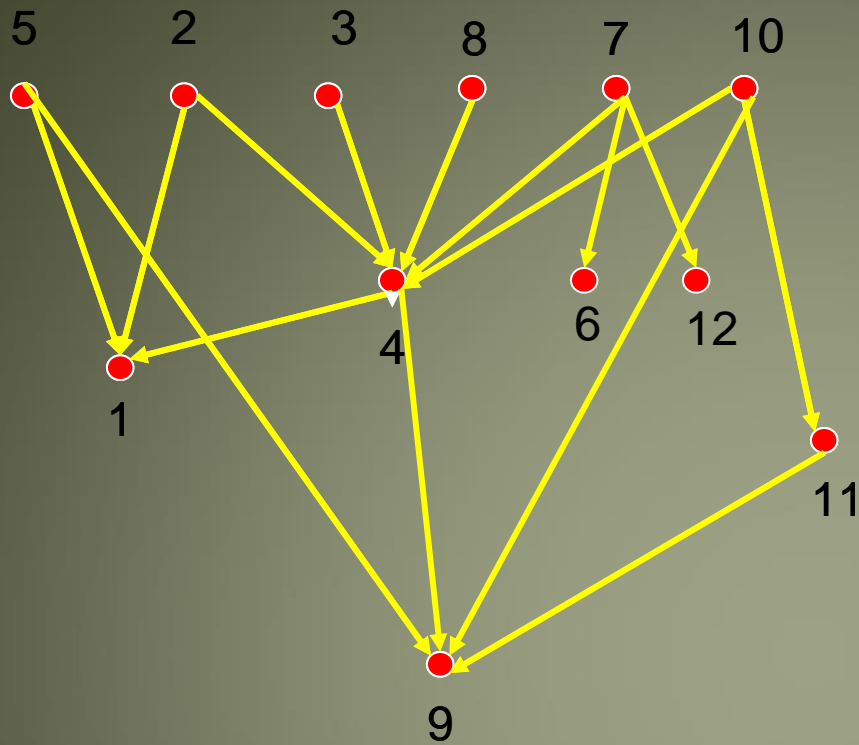What is the boxicity of the competition graph?

Competition graph

1 —————— 4

3 ———— 5
 \   /
  \   /
   2

6   7

Strait of Georgia, British Columbia, Canada

This is an interval graph. Thus, its boxicity is 1.

Competition graph

Strait of Georgia, British Columbia, Canada

32

Malaysian Rain Forest
Due to Harrison
From Cohen, *Food Webs and Niche Space*

Key
1. Canopy – leaves, fruits, flowers
2. Canopy animals – birds, bats, etc.
3. Upper air animals – insectivores
4. Insects
5. Large ground animals – large mammals & birds
6. Trunk, fruit, flowers
7. Middle-zone scansorial animals
8. Middle-zone flying animals
9. Ground – roots, fallen fruit, leaves
10. Small ground animals
11. Fungi

33

Malaysian Rain Forest

Competition Graph

34
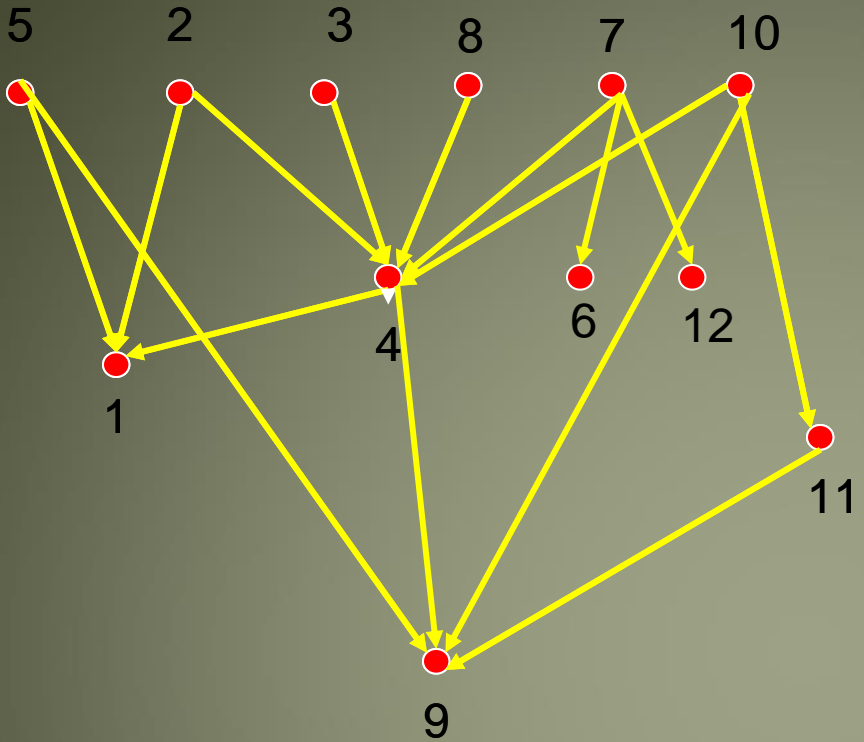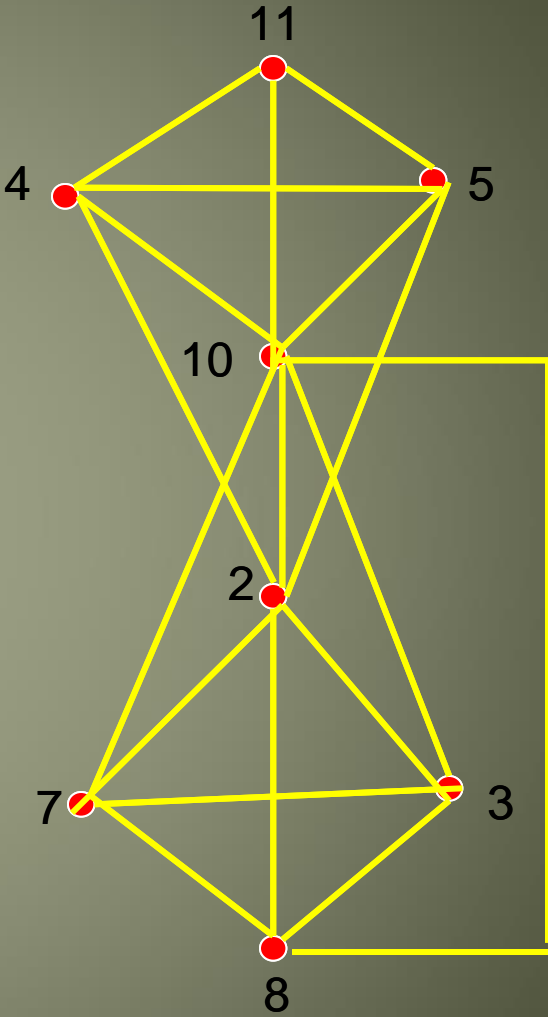
What is the boxicity of the competition graph?

Competition Graph

Malaysian Rain Forest
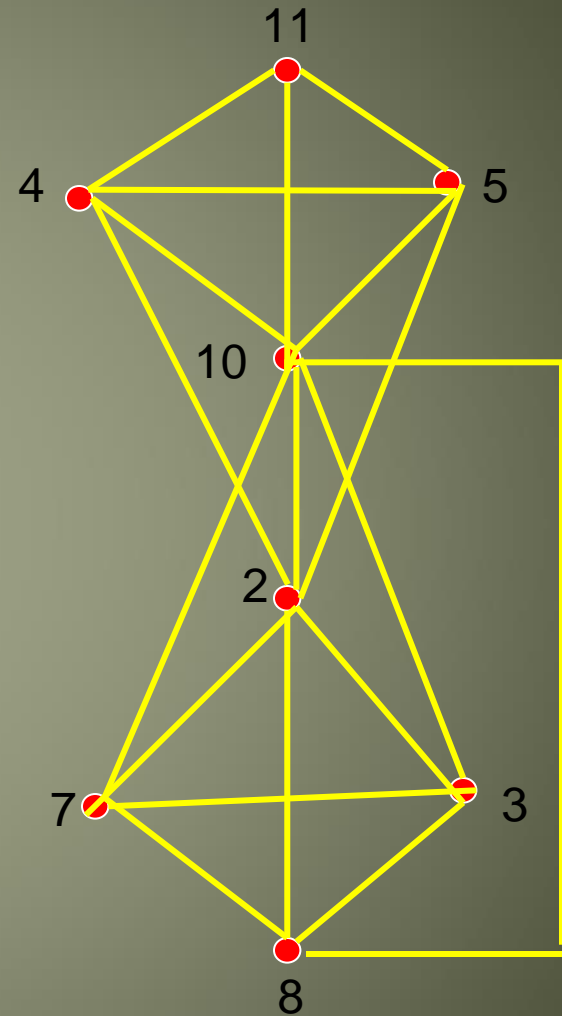
Competition Graph

This is an interval graph. Thus, its boxicity is 1.

11

4

5

2   7

8

3

10

Malaysian Rain Forest

11

4

5

10

2

7

3

8

1   6   9

1   6   9
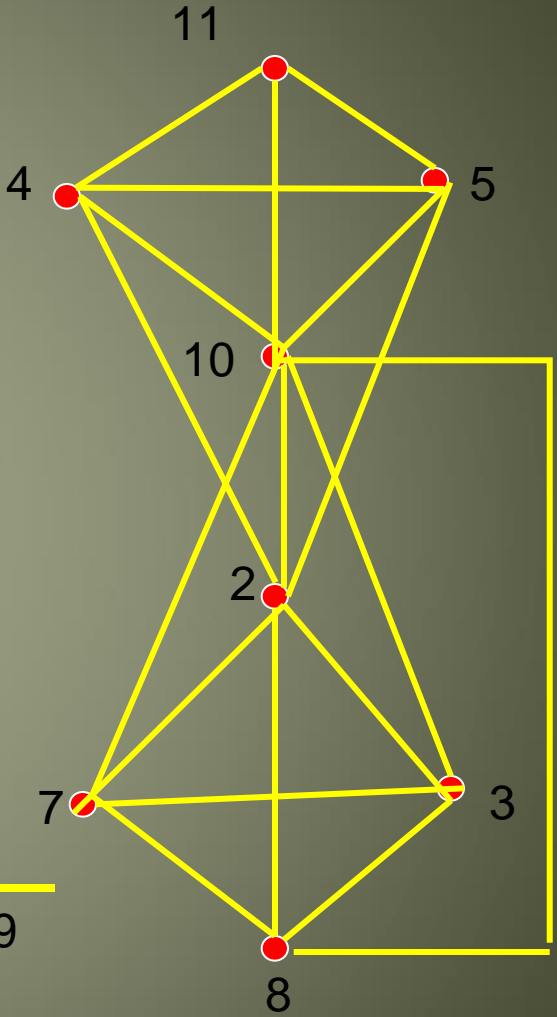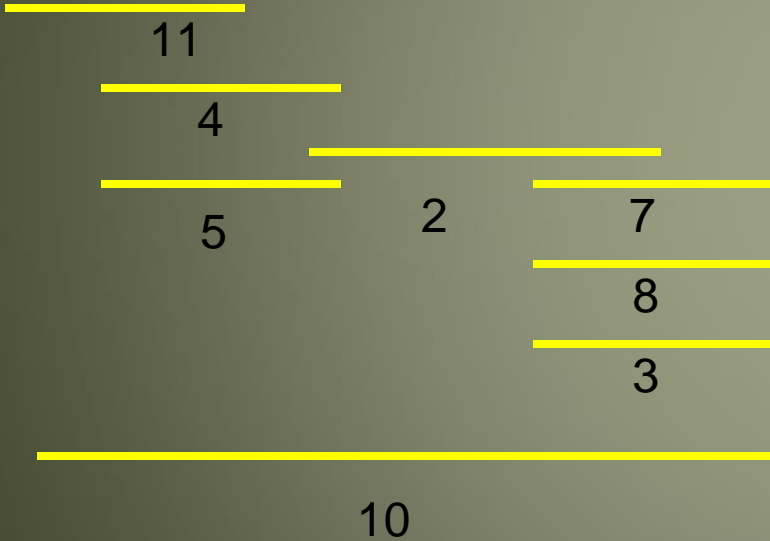
Structure of Competition Graphs

•In the first 12 years after this problem was introduced, every food web studied was found to have a competition graph that was an interval graph.

•In 1976, a Rutgers undergraduate, Gordon Kruse, found the first example of a food web whose competition graph was not an interval graph.
It arose from a complex set of habitats.

•Generally: Food webs arising from "single habitat ecosystems" (homogeneous ecosystems) have competition graphs that are interval graphs.

Structure of Competition Graphs

The remarkable empirical observation of Cohen's that real-world competition graphs are usually interval graphs has led to a great deal of research on the structure of competition graphs and on the relation between the structure of digraphs and their corresponding competition graphs, with some very useful insights obtained.

Competition graphs of many kinds of digraphs have been studied.

In most of the applications of interest, the digraphs studied are acyclic.

Structure of Competition Graphs

- *Statistical Explanations*:
  - Develop models for randomly generating food webs
  - Calculate probability that the corresponding competition graph is an interval graph
  - Much of Cohen's *Food Webs and Niche Space* takes this approach.
  - Cascade model developed by Cohen, Newman, and Briand. But Cohen and Palka showed that under this model, the probability that a competition graph is an interval graph goes to 0 as the number of species increases.

Structure of Competition Graphs

- *Graph-theoretical Explanations:*
  - Analyze the properties of competition graphs that arise from different kinds of digraphs.
  - Characterize the digraphs whose corresponding competition graphs are interval graphs.
  - Much known about the former problem.
  - Latter problem remains the fundamental open problem in the subject.
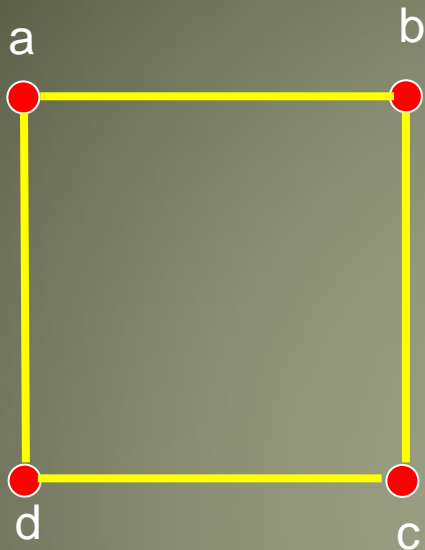
The Competition Number

Suppose D is an acyclic digraph. Then its competition graph must have an isolated vertex (a vertex with no neighbors).

Theorem: If G is any graph, adding sufficiently many isolated vertices produces the competition graph of some acyclic digraph.
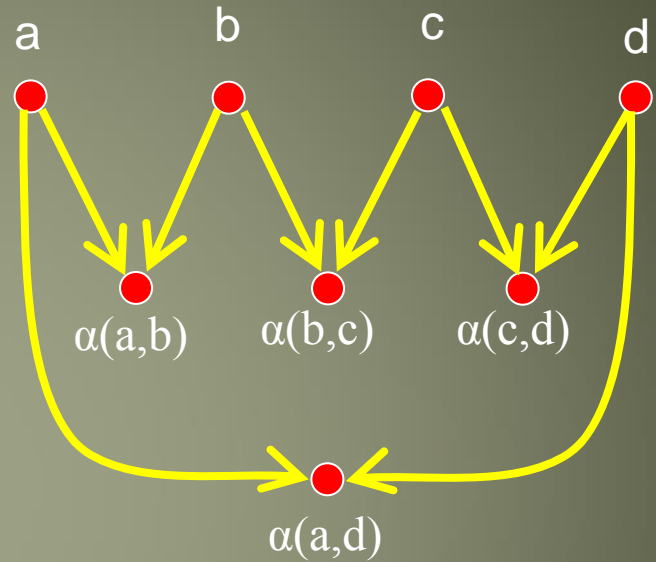
Proof: Construct acyclic digraph D as follows. Start with all vertices of G. For each edge {x,y} in G, add a vertex α(x,y) and arcs from x and y to α(x,y). Then G together with the isolated vertices α(x,y) is the competition graph of D.

The Competition Number

$G = C_4$
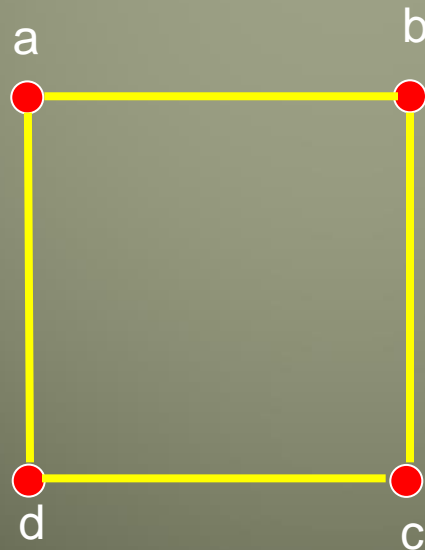
D

$\alpha(a,b)$

$\alpha(b,c)$

$\alpha(c,d)$

$\alpha(a,d)$

$C(D) = G \cup I_4$

● $\alpha(a,b)$

● $\alpha(b,c)$

● $\alpha(c,d)$

● $\alpha(a,d)$

The Competition Number

- Thus, D as shown in previous slide has a competition graph that is not an interval graph.
- In fact, there are examples of competition graphs of acyclic digraphs that have arbitrarily high boxicity.
  - Just start with a graph of boxicity b.
  - Add sufficiently many isolated vertices to make the graph into a competition graph. (Adding isolated vertices does not change the boxicity.)
- *Thus, the empirical observations tracing back to Joel Cohen are truly surprising.*

43

The Competition Number

If G is any graph, let k be the smallest number so that $G \cup I_k$ is a competition graph of some acyclic digraph.

$k = k(G)$ is well defined.

It is called the ***competition number*** of G.

Thus $k(C_4) = 4$ from the previous example

The Competition Number

Competition numbers and boxicity are known for many interesting graphs and classes of graphs.

However:

<u>Theorem (Cozzens 1980)</u>: It is NP-complete to determine the boxicity of a graph, even if a graph has boxicity greater than or equal 2.

<u>Theorem (Opsut 1982):</u> It is an NP-complete problem to compute k(G).

The Competition Number

•Characterization of which graphs arise as competition graphs of acyclic digraphs comes down to the question: Given a graph, how many isolated vertices is it necessary to add to make it into a competition graph?

•There has been extensive work over the years on algorithms for calculating the competition number.

Data Gathering: Community Food Webs, Sink Food Webs, Source Food Webs

- How one gathers data about food webs can influence your conclusions about competition graphs, biodiversity, etc.
- A community food web includes all predation relations among species.
- In practice, we don't always get all this data. We might start with some species, look for species they prey on, look for species they prey on, etc.
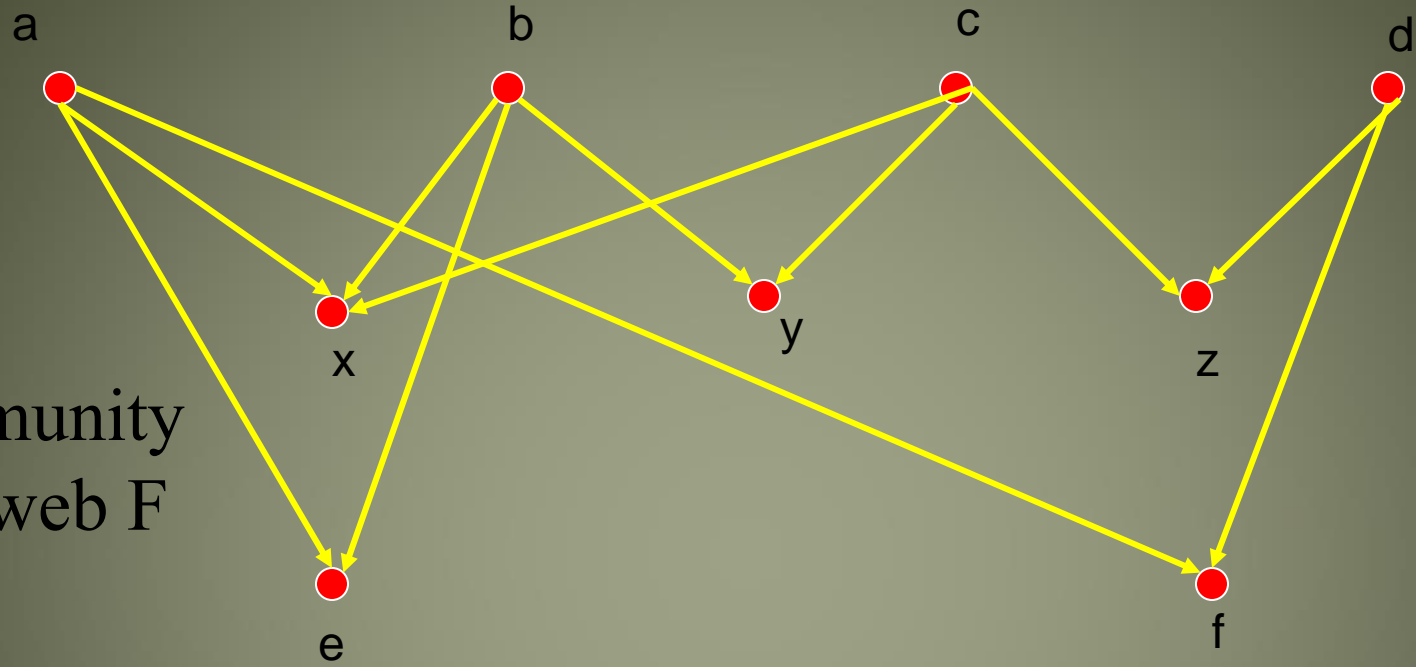
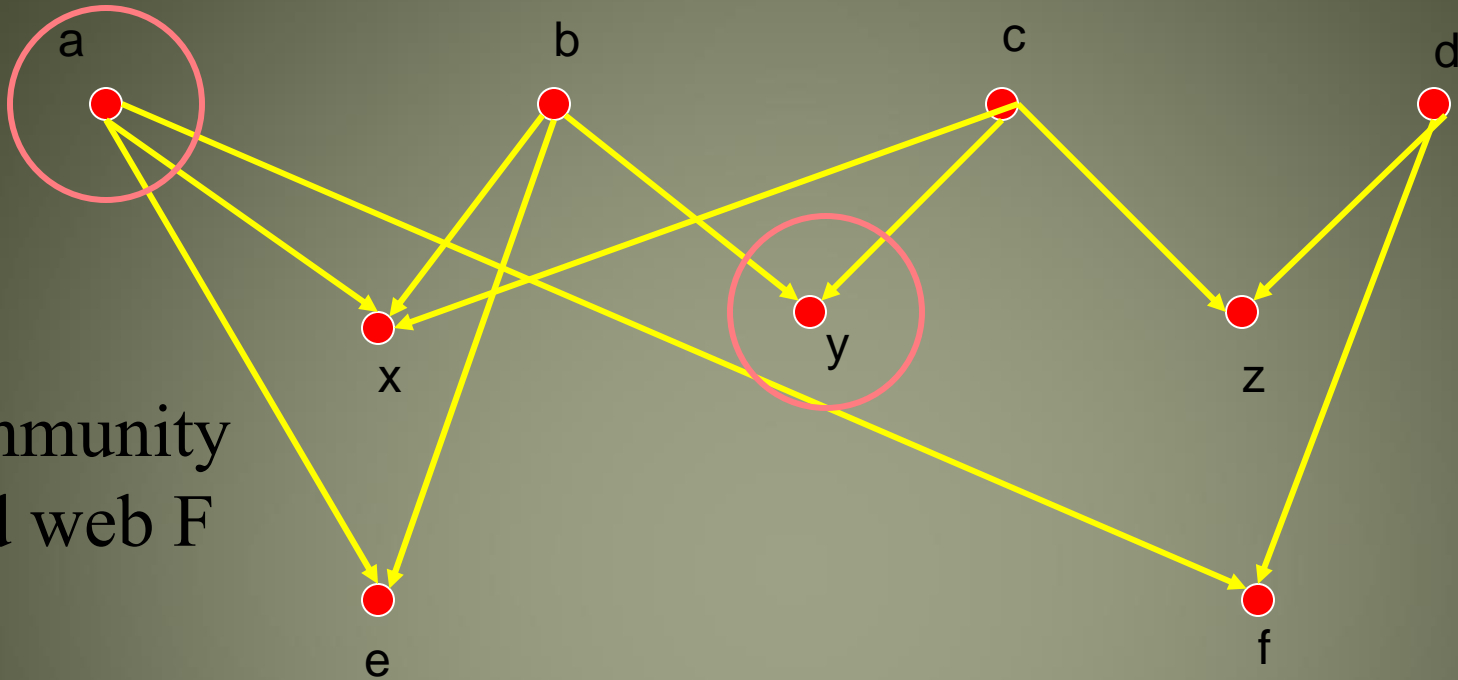# Data Gathering: Community Food Webs, Sink Food Webs, Source Food Webs

- Suppose F is a community food web.
- Let W be a set of species in F (ones we start with).
- Let X be the set of all species that are reachable by a path in F from vertices in W.
  - So, we start with vertex of W, find its prey, find prey of the prey, etc.
- Let Y be the set of all species that reach vertices of W by a path in F.
  - So we start with vertex of W, find its predators, find predators of those predators, etc.

Data Gathering: Community Food Webs, Sink Food Webs, Source Food Webs

- Suppose F is a community food web.
- Let W be a set of species in F (ones we start with).
- Let X be the set of all species that are reachable by a path in F from vertices in W.
  - The subgraph induced by vertices of X is called the *sink food web* corresponding to W.
- Let Y be the set of all species that reach vertices of W by a path in F.
  - The subgraph induced by vertices of Y is called the *source food web* corresponding to W.

Data Gathering
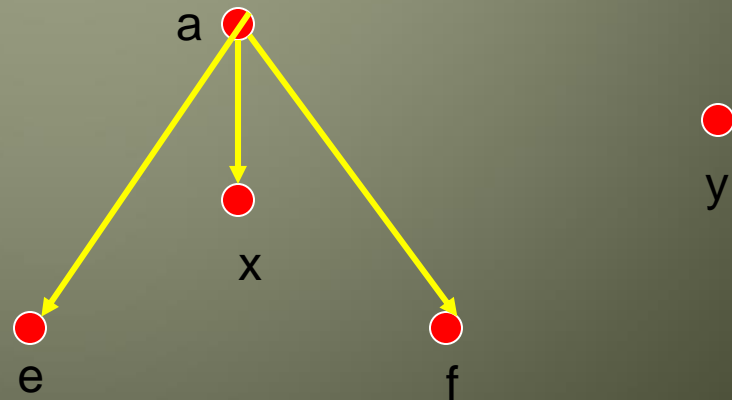

Community food web F
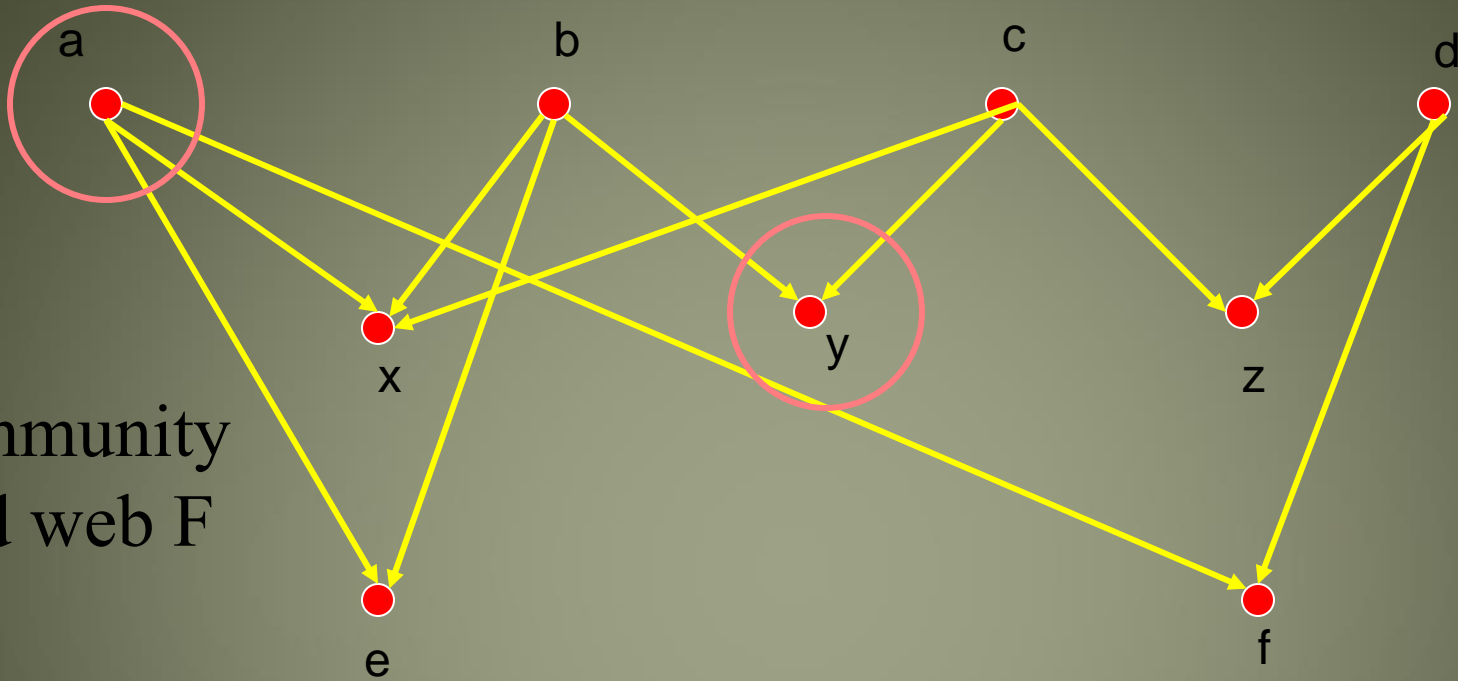
Data Gathering

Community food web F

W = {a,y}
What is the sink food web?

X = {a,x,e,f,y}

Sink food web from W

Data Gathering

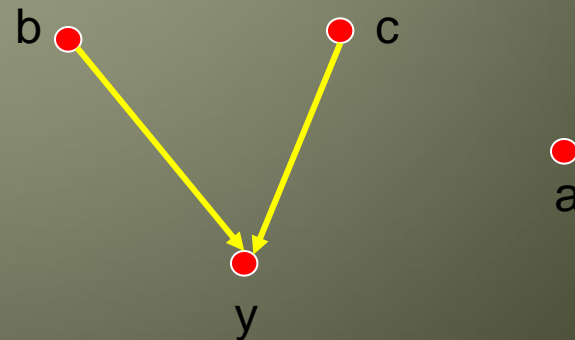Community food web F
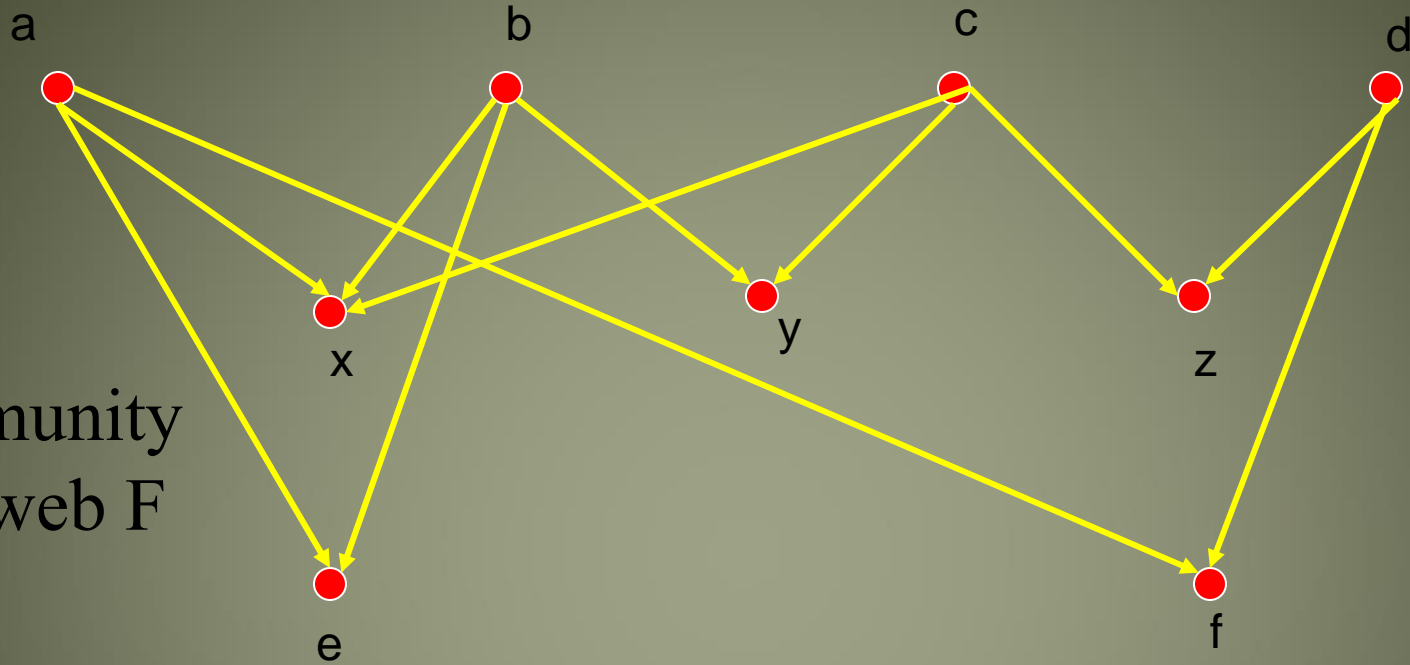
Source food web from W
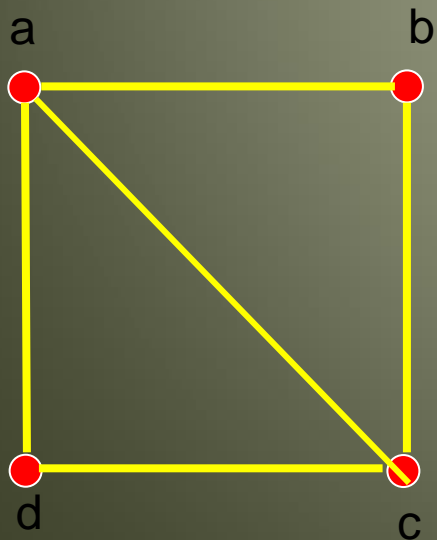
W = {a,y}
What is the source food web?
Y = {a,b,c,y}

Data Gathering: Community Food Webs, Sink Food Webs, Source Food Webs

•Theorem (Cohen): A community food web has a competition graph that is an interval graph if and only if every sink food web contained in it does.

•However: A community food web can have a competition graph that is an interval graph while some source food web contained in it has a competition graph that is not an interval graph.
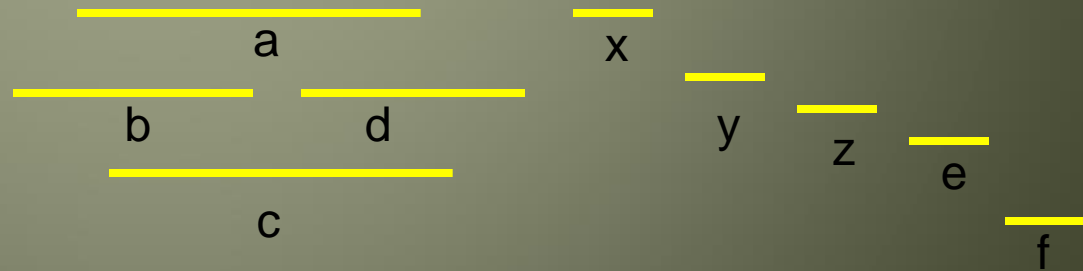
Data Gathering

a b c d

x

y z

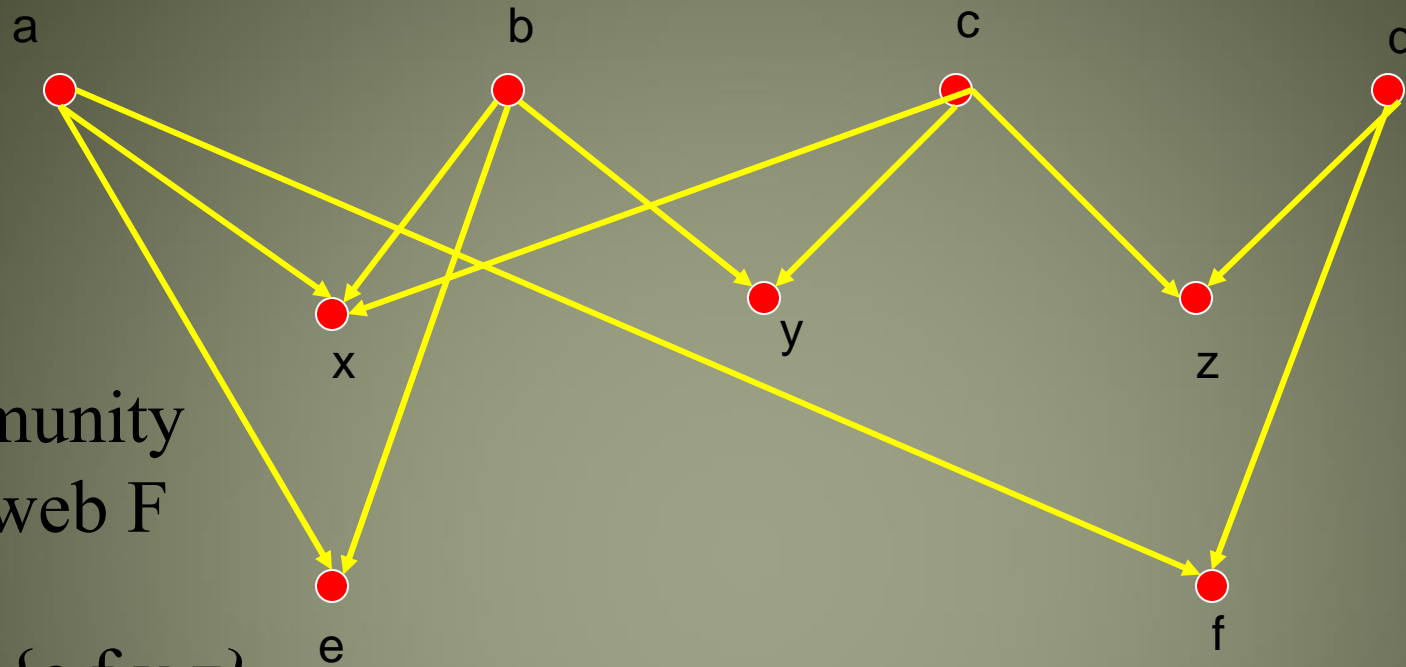Community food web F

e

f

a b

- x
- y
- z
- e
- f

a

b d

c

x

y

z

e

f

The competition graph of F is an interval graph.

Data Gathering

a   b   c   d

x   y   z

Community
food web F

e   f
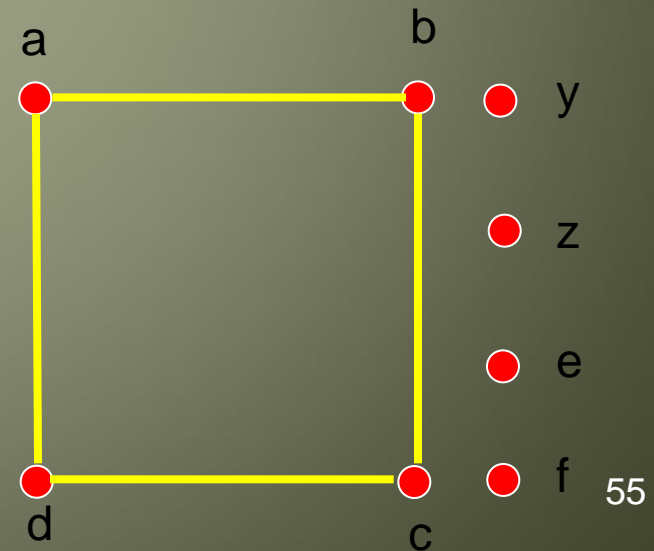
$W = \{e,f,y,z\}$

What is the competition graph
of the source food web of W?

$Y = \{a,b,c,d,e,f,y,z\}$

Competition graph of the
source food web from W

This is not an interval graph.

a   b   y

z

e

d   c   f

Data Gathering: Community Food Webs, Sink Food Webs, Source Food Webs

•This surprising result points up some of the difficulties involved in understanding the structure of competition graphs.

•It also leads to interesting caveats about general conclusions using models that are tested with data.

The Interval Graph Competition Graph Problem

- It remains a challenge (dating back to 1968) to understand what properties of food webs give rise to competition graphs of boxicity 1, i.e., interval graphs.
- In a computational sense this is easy to answer:
  - Given a digraph, compute its competition graph (easy)
  - Determine if this is an interval graph (well known to be solvable in linear time)

The Interval Graph Competition Graph Problem

•More useful would be results that explain the structural properties of acyclic digraphs that give rise to interval graph competition graphs.

•However, such results might be difficult to find:

•There is no list L (finite or infinite) of digraphs such that an acyclic digraph D has an interval graph competition graph if and only if does not have an induced subgraph in the list L.

Open Problems

- There are, however, results with extra assumptions about the acyclic digraph D.
- Example: It is useful is to place limitations on the indegree and outdegree of vertices (the maximum number of predator species and maximum number of prey species for any given species in the food web). Is this reasonable however? Then there are some results with forbidden lists L. (e.g., Hefner, et al., 1991).

Biodiversity

- What are the implications of this work for the measurement of biodiversity?
- *General Challenge: Can we "derive" "dimensions" of biodiversity from other important properties of ecosystems?*