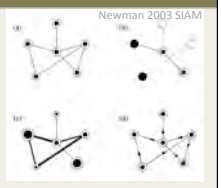
Structural Network Analysis

Stuart R. Borrett
University of North Carolina Wilmington

Network Structure Fundamentals

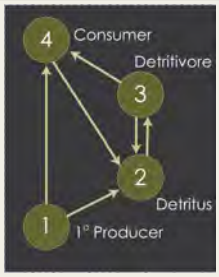


$G = \{V, E\}$

- What type of graph?
- What do the vertices and edges represent?
- How many vertices (nodes) – n
- How many edges (links) – L
- Connectance – $C = L/n^2$ or $L/n(n-1)$

Does not describe any pattern of connections

Pathways



Pathway = a sequence of edges

$1 \rightarrow 2 \rightarrow 3 \rightarrow 4 \rightarrow 2 \rightarrow 3$

Pathway length is the number of edges in the sequence (5 in example)

Distance (Geodesic) is the minimum path length required to get from one node to another.

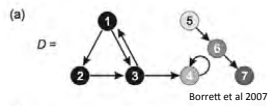
Find with Dijkstra's algorithm.

$$\mathbf{D} = \begin{bmatrix} 0 & \infty & \infty & \infty \\ 1 & 0 & 1 & 1 \\ 2 & 1 & 0 & 2 \\ 1 & 2 & 1 & 0 \end{bmatrix}$$

Reachability

Diameter of a graph is the *mean* or *maximum* distance

Components



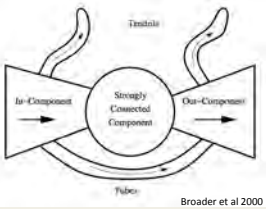
Borrett et al 2007

In a *component* the nodes are reachable across a pathway ignoring direction

In an *SCC* its possible to move from any node to any node over a pathway of some length following directions.

(a) Is an example digraph composed of 2 components, but 5 strongly connected components

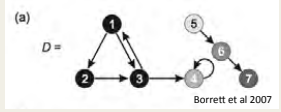
Structure of the Internet



Broader et al 2000

What are the 5 SCC's?

Components



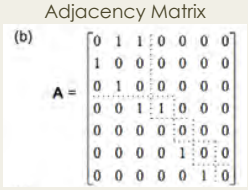
Borrett et al 2007

In a *component* the nodes are reachable across a pathway ignoring direction

In an *SCC* its possible to move from any node to any node over a pathway of some length following directions.

(a) Is an example digraph composed of 2 components, but 5 strongly connected components

Adjacency Matrix



Nodes ordered to in "block diagonal form"

What are the 5 SCC's?

Food Web Modularity

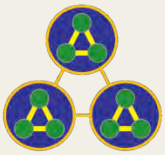
Stuart R. Borrett, Brian D. Fath, & Bernard C. Patten

2007, Journal of Theoretical Biology

Food Web Modularity

Hypothesis:

Food webs are comprised of loosely connected subunits



Rational

Ease of assembly (Simon 1962)

Stability (May 1974)

Evidence

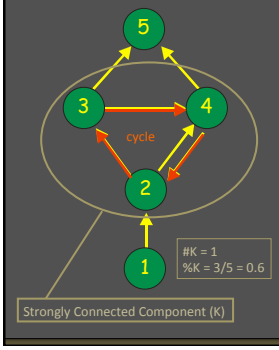
Uncommon (Pimm & Lawton 1980)

No cycles (Cohen et al. 1990)

Cohesive Subgroups (Krause et al. 2003)

Inconclusive results & multiple methods

Modules = Strongly Connected Components



Possible to reach every node from every other via pathways of some length

Ecological Significance

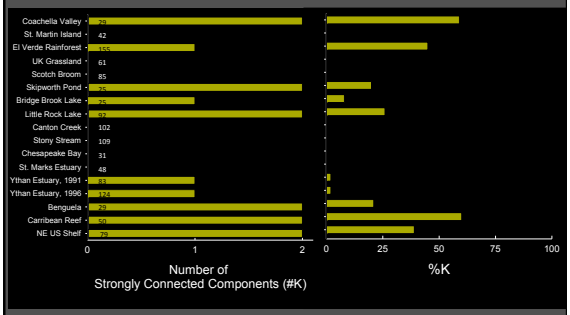
- indirect effects
- autocatalytic
  - systems that catalyze their own production
- indirect mutualism

Food Web Structure†

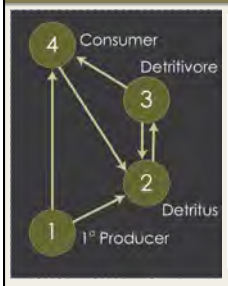
Habitat	Food Web	Taxa	n	C	%I
Terrestrial	Coachella Valley	30	29	0.31	90
	St. Martin Island	44	42	0.12	69
	El Verde Rainforest	156	155	0.06	69
	UK Grassland	75	61	0.03	69
	Scotch Broom	154	85	0.03	40
Lake/Pond	Skipworth Pond	35	25	0.32	92
	Bridge Brook Lake	75	25	0.17	68
	Little Rock Lake	181	92	0.12	86
Stream	Canton Creek	108	102	0.07	22
	Stony Stream	112	109	0.07	27
Estuary	Chesapeake Bay	33	31	0.07	52
	St. Marks Estuary	48	48	0.10	80
	Ythan Estuary, 1991	92	83	0.06	54
	Ythan Estuary, 1996	134	124	0.04	56
Marine	Benguela	29	29	0.24	93
	Carribean Reef, small	50	50	0.22	94
	NE US Shelf	81	79	0.22	94

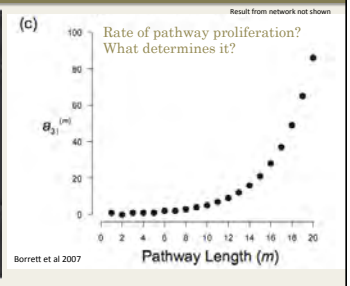
† Previously reported by Williams & Martinez 2000; Dunne et al. 2002, 2004

Modularity – Strongly Connected Components



Pathway Enumeration & Proliferation





Input $\mathbf{B} = \underbrace{\mathbf{I}}_{\text{Boundary}} + \underbrace{\mathbf{A}^1}_{\text{Direct}} + \underbrace{\mathbf{A}^2 + \mathbf{A}^3 + \ldots + \mathbf{A}^m + \ldots}_{\text{Indirect}}$

Output $\mathbf{B}' = \underbrace{\mathbf{I}}_{\text{Boundary}} + \underbrace{\mathbf{A}'^1}_{\text{Direct}} + \underbrace{\mathbf{A}'^2 + \mathbf{A}'^3 + \ldots + \mathbf{A}'^m + \ldots}_{\text{Indirect}}$

Conclusions

- Food webs can have modular form
  - not as much as we might expect given the stability or assembly hypotheses
- #K is not correlated with n
- Functional significance
  - Cycles distribute indirect effects

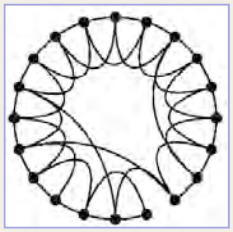
Other Patterns

Small Worlds

Compared to randomly constructed networks (RG)

1) node clustering is larger
2) maximum distance is lower than expected

Watts & Strogatz 1998



Examples

Stanly Milgram's letter experiment

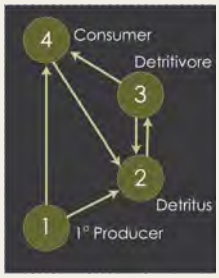
6 degrees of {Kevin Bacon}

Consequence

Movement in the network is faster than in RG with same {n,C}

Degree Distribution -- Centrality

Node degree is the number of edges incident to a node



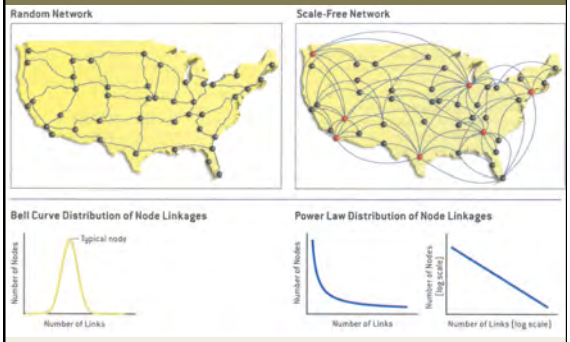
In	Out	Total
0	2	2
3	1	4
1	2	3
2	1	3

Centrality

which node is the most central?

Many ways of determining …

Scale Free Networks



Barabasi and Bonabeau 2003

Robustness to Node Deletion

Consequences of scale free pattern include

1) robustness to random attacks (node deletion)
2) increased sensitivity to targeted attacks

Albert and Barabasi demonstrated this for the internet.

Dunne et al. 2002 Found that few food webs exhibit the scale-free distributions, but the distributions are not Poisson either.

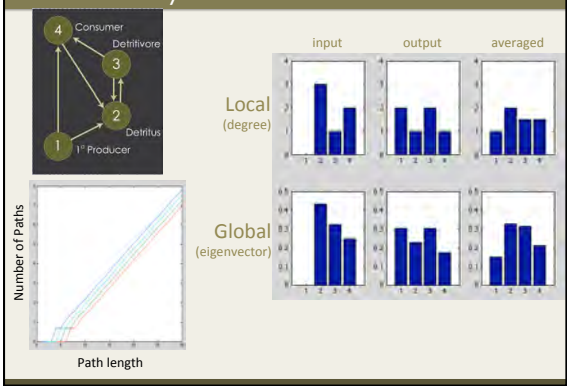
Examples of Scale Free Network

NETWORK	NODES	LINKS
Cellular metabolism	Molecules involved in burning food for energy	Participation in the same biochemical reaction
Hollywood	Actors	Appearance in the same movie
Internet	Routers	Optical and other physical connections
Protein regulatory network	Proteins that help to regulate a cell's activities	Interactions among proteins
Research collaborations	Scientists	Co-authorship of papers
Sexual relationships	People	Sexual contact
World Wide Web	Web pages	URLs

Barabasi and Bonabeau 2003

How general a phenomenon in complex systems?
What processes generate this patterns?

Centrality: Local vs. Global Walks



Other Topics

- Network Motifs
  - Frequency of smaller patterns
- Mixing Patterns/ Assortativity
  - Node types by pattern
- Community Structure (modularity)
  - Clustering Analysis
  - Hierarchical Clustering
- Betweeness Centrality
- Generative Models
  - Infer rules to grow networks with given patterns
  - Preferential Attachment Algorithm
- Graph Layout Algorithms

Suggested Activities

Each Person/Team Should Select a Model
Use enaR to complete the activities

Activity: Network Structural Properties

- Load oyster reef model & perform structural analysis
  - load(enaR)
  - data(oyster)
  - S=enaStructure(oyster)
- Extract adjacency matrix
  - A= S$A
- Network Statistics
  - S$ns

Activity: Pathways Enumeration

- Load the Oyster Reef Model & get A
- Calculate A^2 - mExp(A,2)
- Calculate A^3
  - Can you identify all 4 pathways from node 2 to itself of length 3?
- Use a for-loop to get the data to plot the relationship between path length and the number of paths.