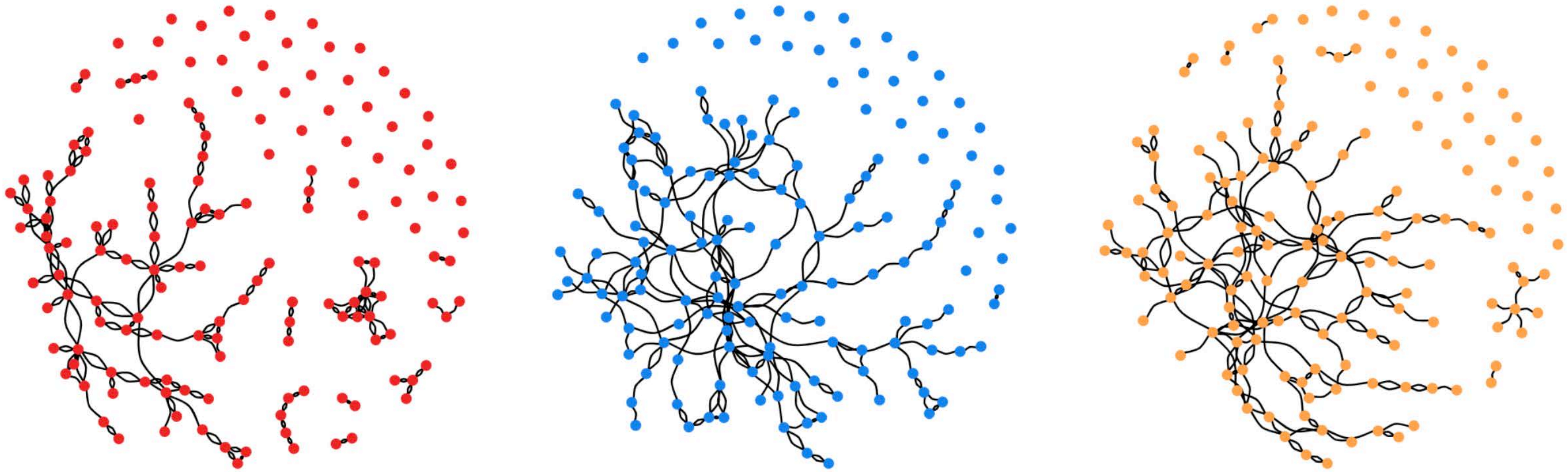
Network structure

Analysis of Biological-Ecological Networks 2026

Shai Pilosof

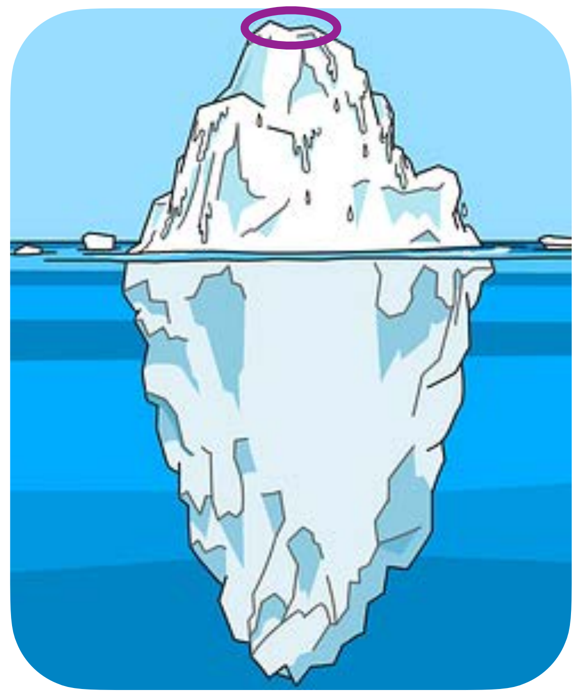


www.ecomplab.com
pilos@post.bgu.ac.il

Ben-Gurion University of the Negev

Class goals

1. Define networks.
2. Introduce basic metrics of structure, focusing on the node and clique levels*.
3. Relate structure to ecological processes and functions.



* This is not even the beginning of the tip of the iceberg. **You must consult papers and books.**

Outline

- Classes of networks
- Density and connectance
- Degree / degree distributions
- Motifs
- Clustering coefficient
- Centrality
- Network comparisons

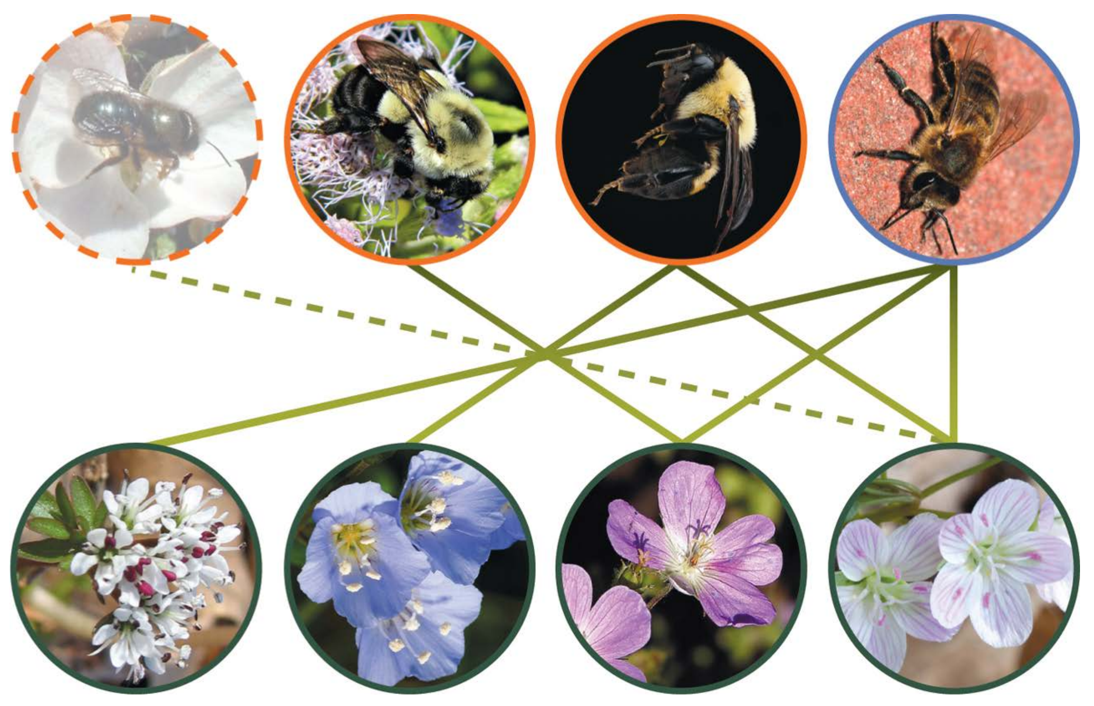
This will follow us throughout the course.

1. Quantify the ***structures*** of biological networks.
2. Relate these structures to ***dynamics, stability*** and ***function***.
3. Understand how and why these structures were formed (***generative processes***).

Microscopic rules

- Interaction constraints
- Evolutionary history
- Node heterogeneity
- Environmental context
- Structure–dynamics

Emergent properties

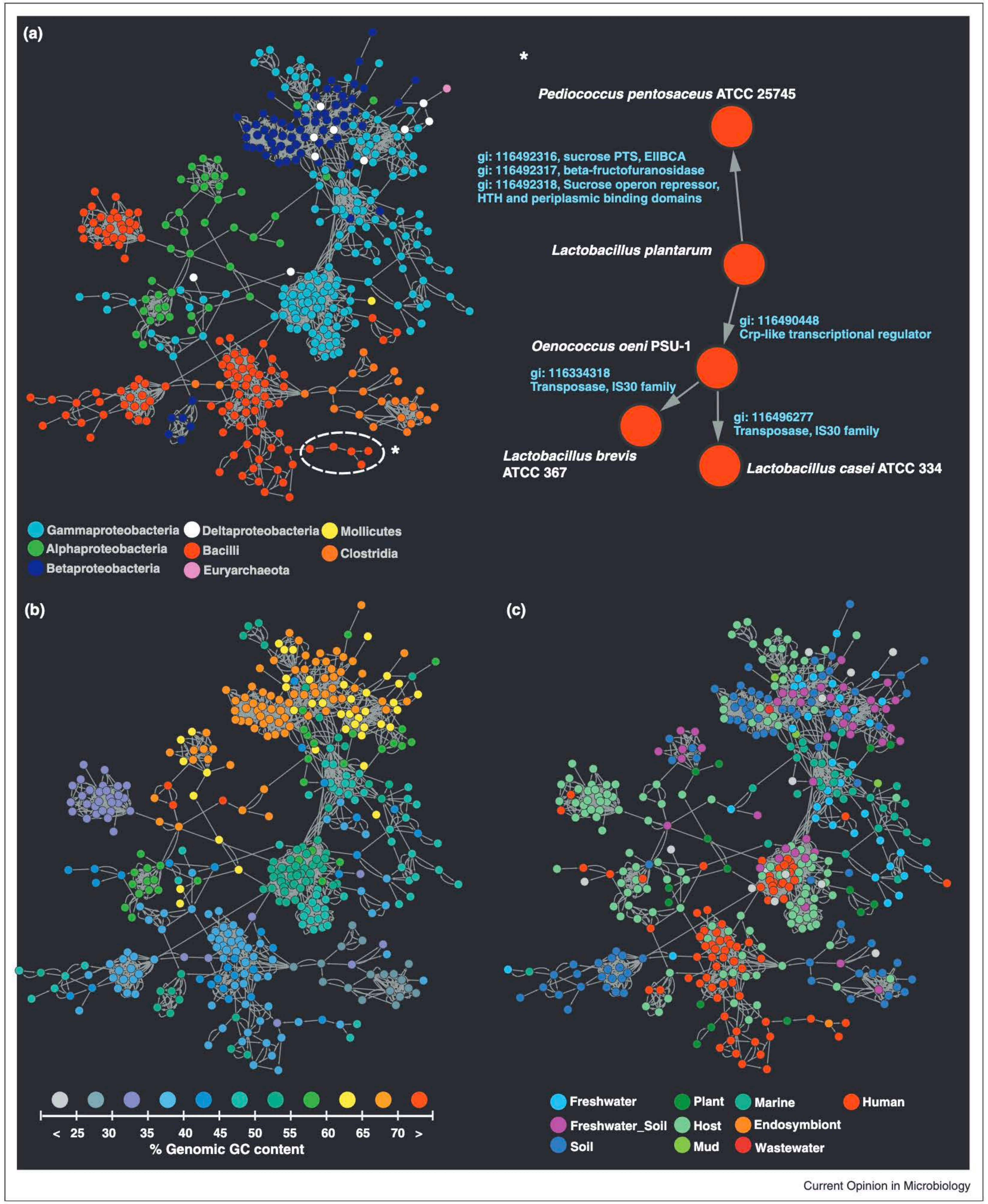


Function

Cell development
Neuro-generation
Pathogen spread
Ecosystem functions
Biological control

Stability

Phage HGT network



Popa et al 2011

Why quantify network structure?

- Describe complexity compactly.
- Link structure to ecological processes.
- Compare systems (detect regularities) across space, time, and scales.
- Generate and test biological hypotheses.

Typical questions involving network structure

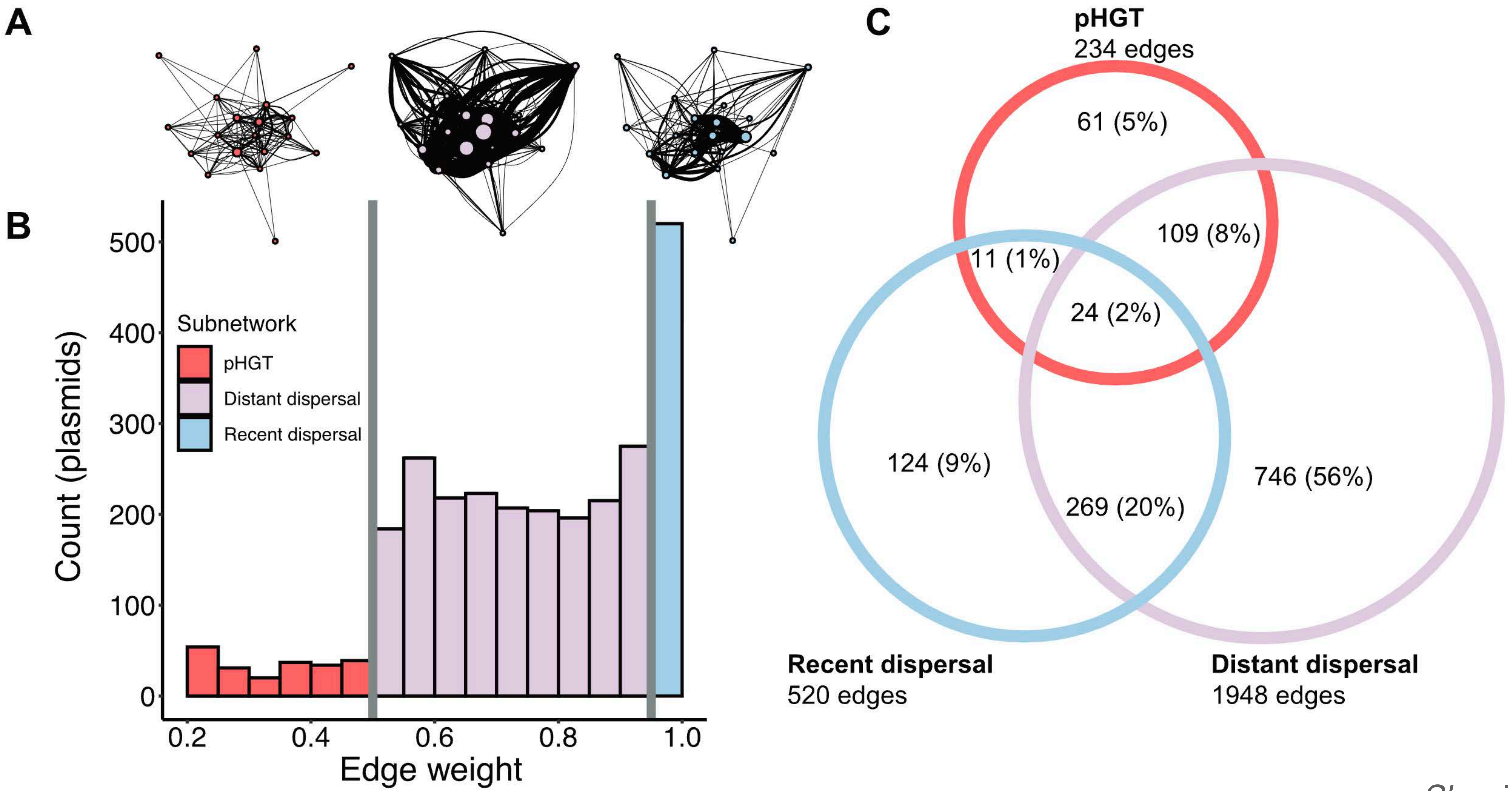
Who interacts with whom, and how uneven are interactions?
(Generality, specialization, heterogeneity)

Which components are central or influential?
(Flow, control, vulnerability)

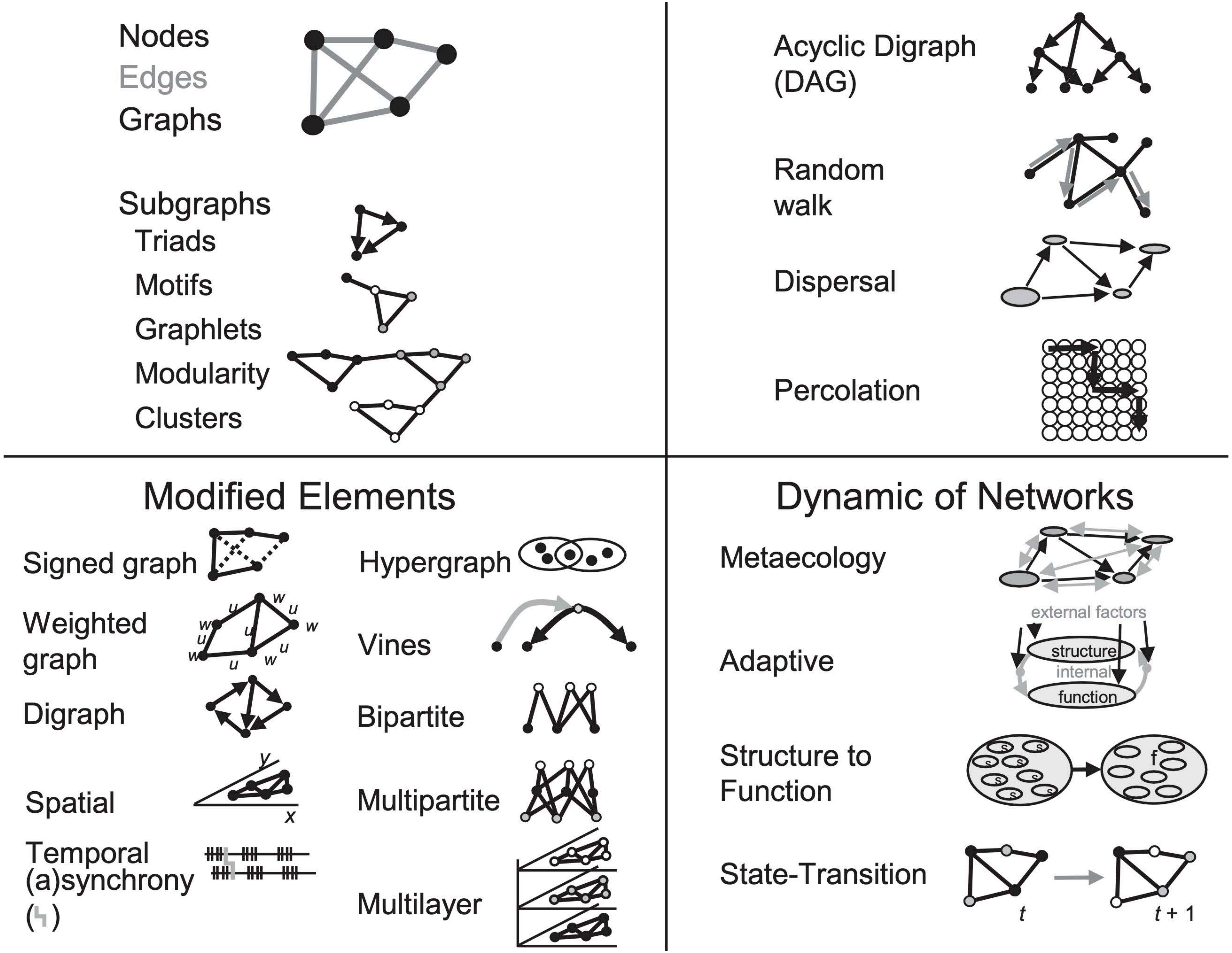
What patterns emerge beyond pairwise interactions?
(Indirect effects, local substructures)

How does structure vary across systems or change under perturbation?
(Space, time, environmental change)

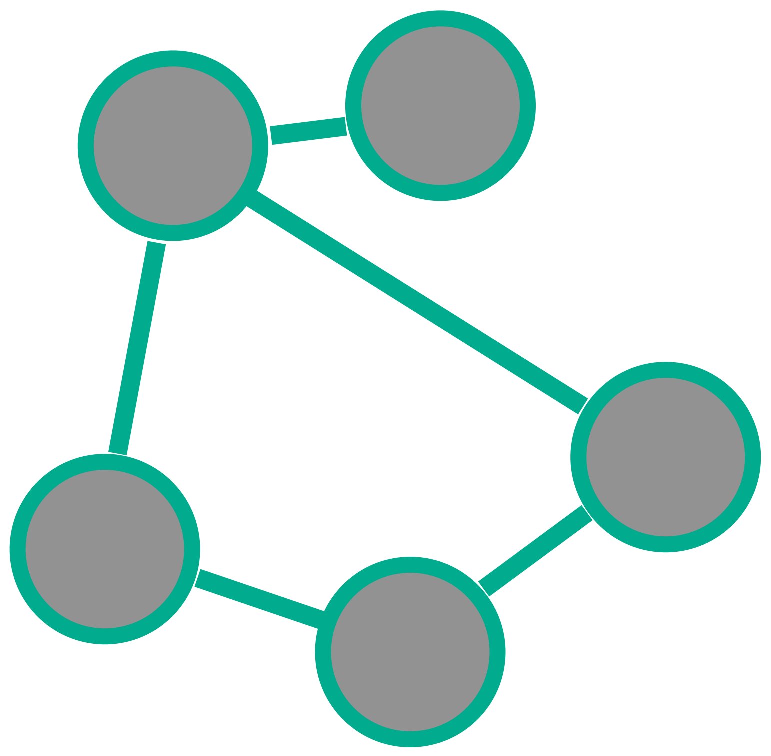
Why quantify network structure?



Shapiro,…,Pilosof 2023

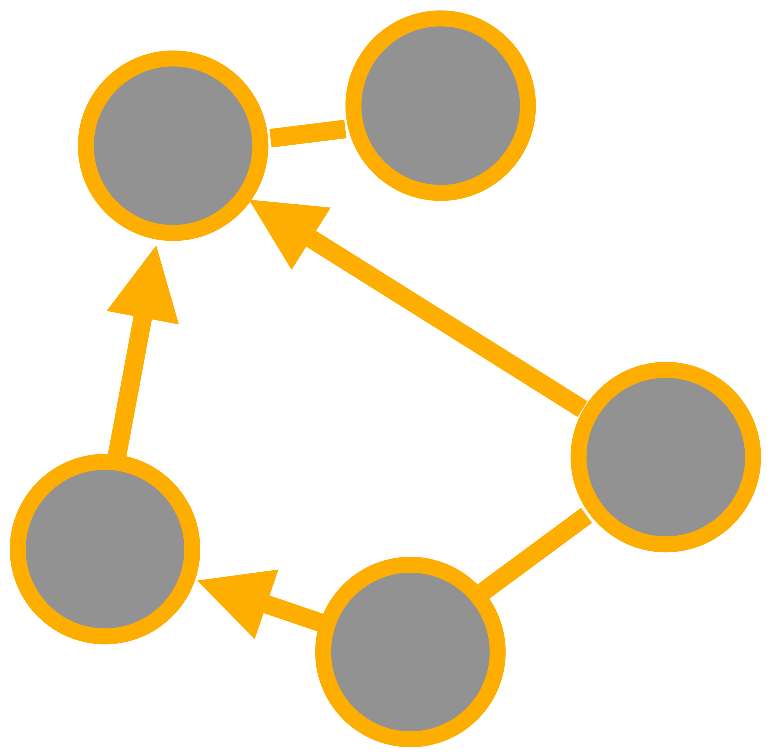


Dale and Fortin 2021



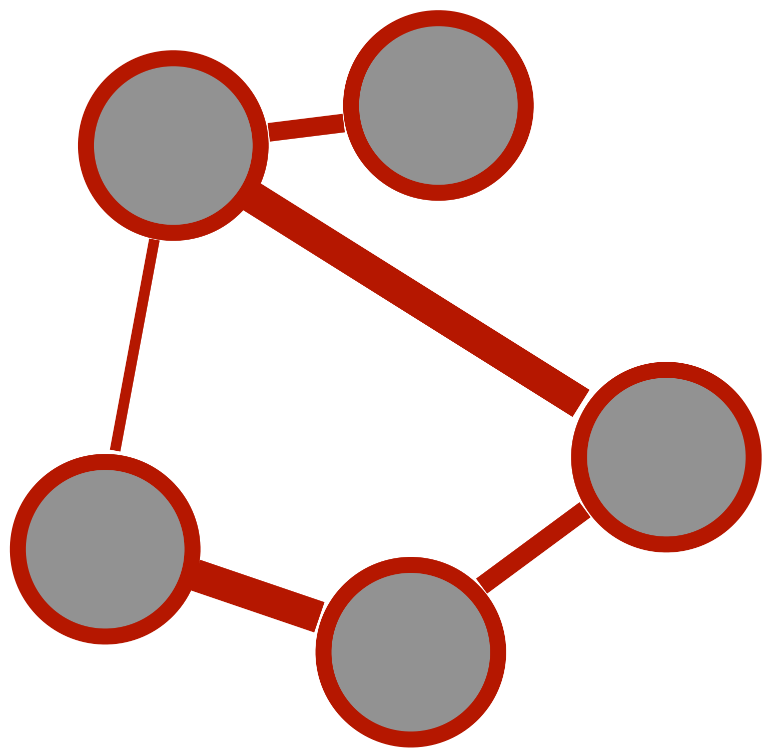
Binary (unweighted)
undirected

$$a_{ij} = a_{ji}$$
$$a_{ij} \in \{0,1\}$$



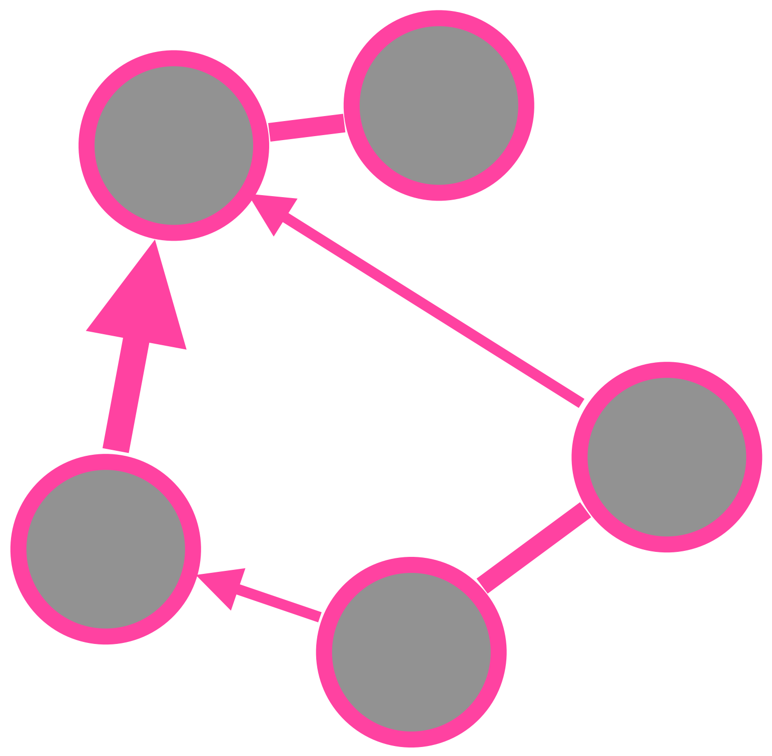
Binary (unweighted)
directed

$$a_{ij} \in \{0,1\}$$



Weighted
undirected

$$a_{ij} = a_{ji}$$



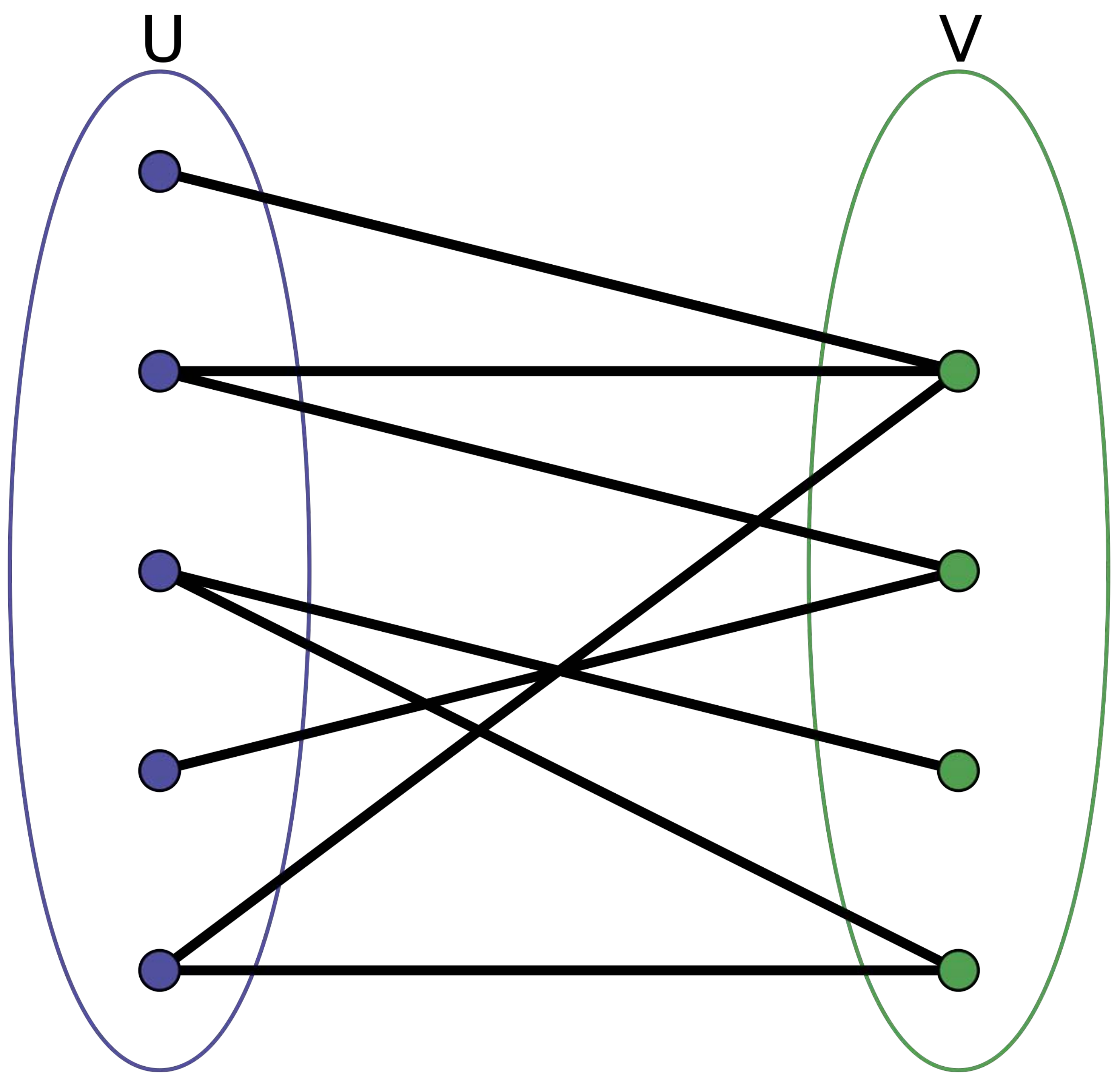
Weighted
directed

$$A = \begin{pmatrix} a_{11} & a_{12} & \cdots & \cdots & a_{1N} \\ a_{21} & a_{22} & \cdots & \cdots & a_{2N} \\ \vdots & \cdots & \ddots & \cdots & \vdots \\ \vdots & \cdots & \cdots & \ddots & a_{MN} \\ a_{N1} & a_{N2} & \cdots & a_{NM} & a_{NN} \end{pmatrix}$$

Adjacency matrix
(Nodes are *adjacent* when they are *incident* to a common edge)

Examples?

$$G = (V, E) \qquad E \subseteq \{\{x, y\} \mid \{x, y\} \in V\}$$



Two disjoint sets of nodes.

Bipartite graphs can be weighted and directed.

$$I = \begin{pmatrix} a_{11} & a_{12} & \cdots & \cdots & a_{1N} \\ a_{21} & a_{22} & \cdots & \cdots & a_{2N} \\ \vdots & \cdots & \ddots & \cdots & \vdots \\ a_{M1} & \cdots & \cdots & \ddots & a_{MN} \end{pmatrix}$$

Incidence matrix with dimensions:

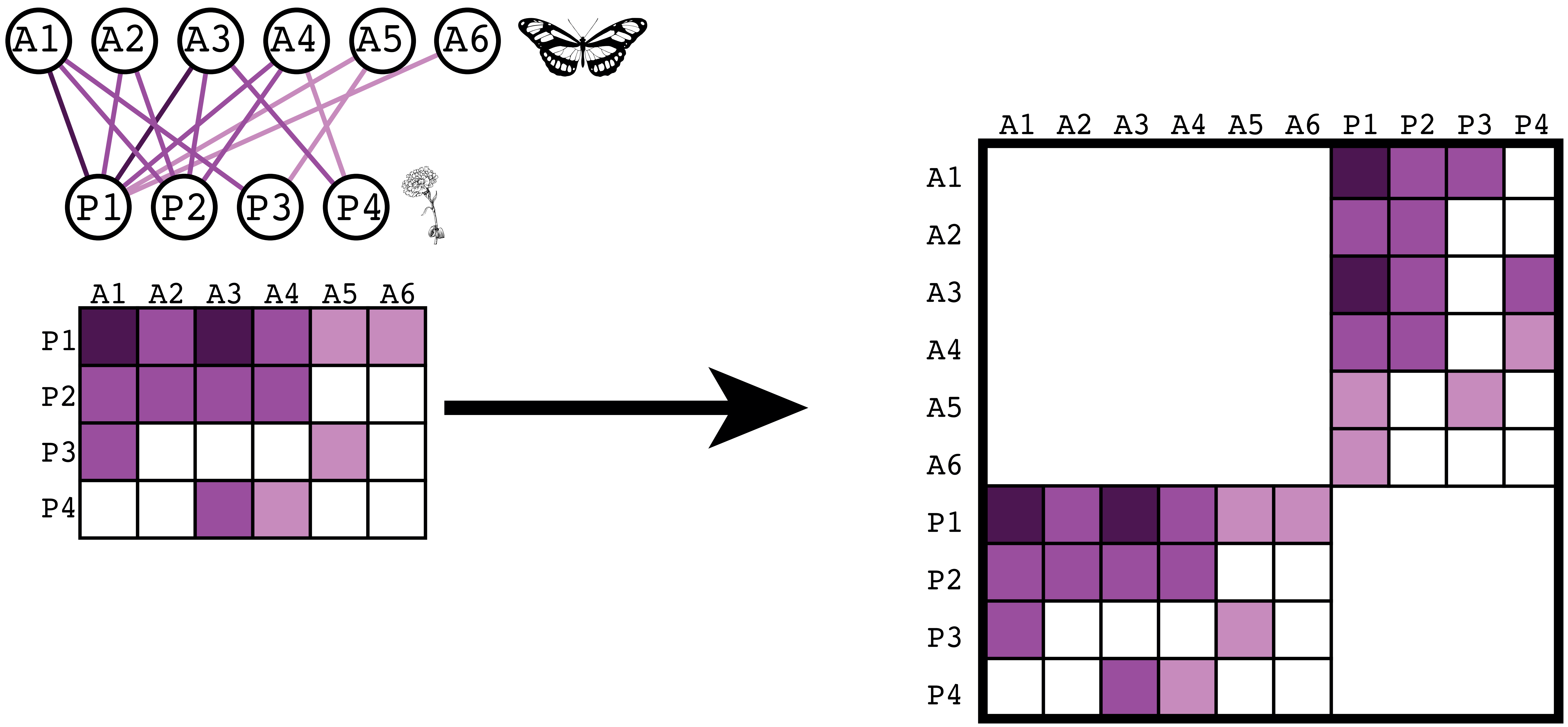
M (rows) and N (columns)

Examples?

$$G = (U, V, E) \quad E \subseteq \{\{x, y\} \,|\, X \in V, y \in U\}$$

A word on incidence and adjacency matrices

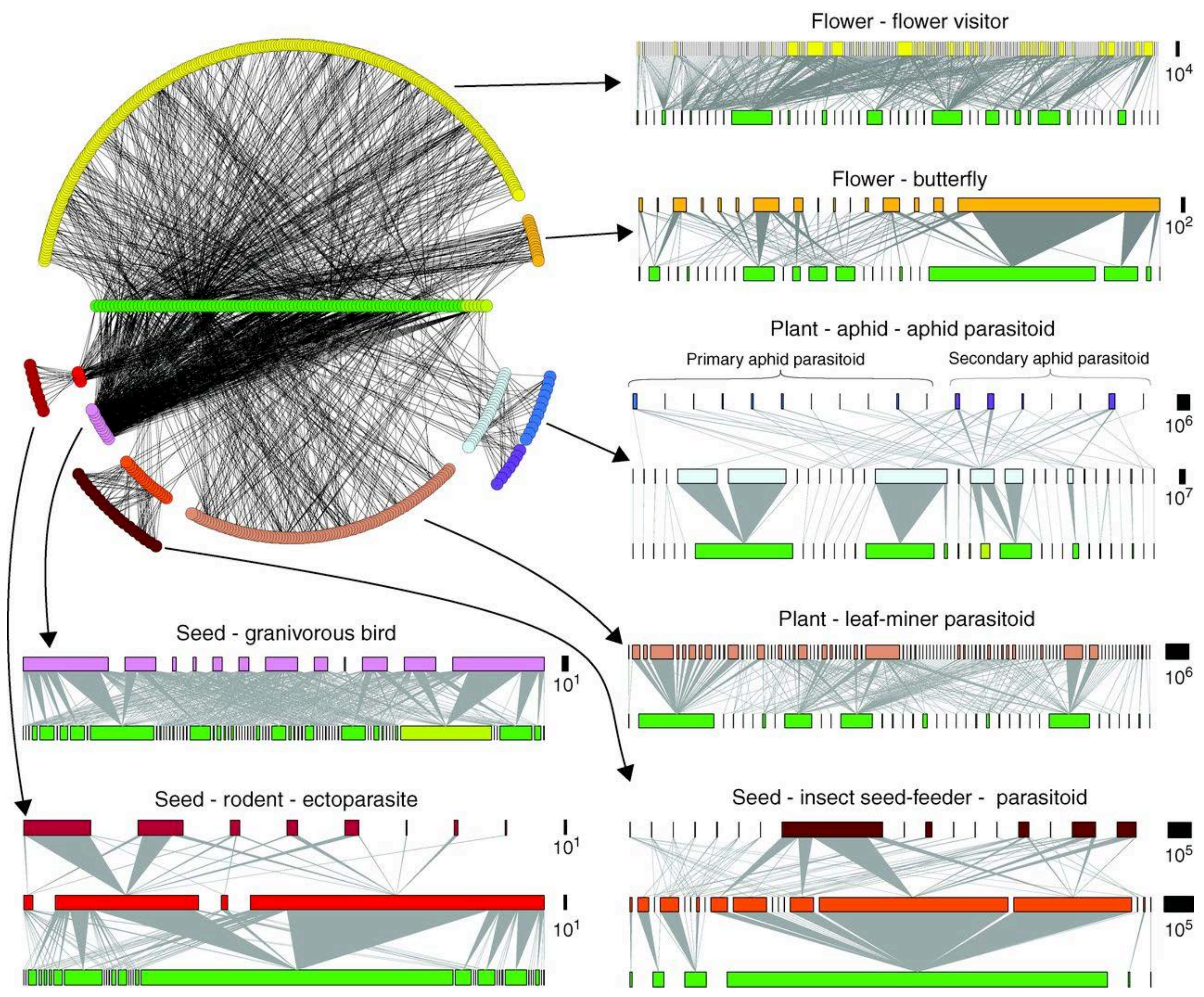
$$\mathscr{A} = \begin{pmatrix} 0 & B^T \\ B & 0 \end{pmatrix}$$

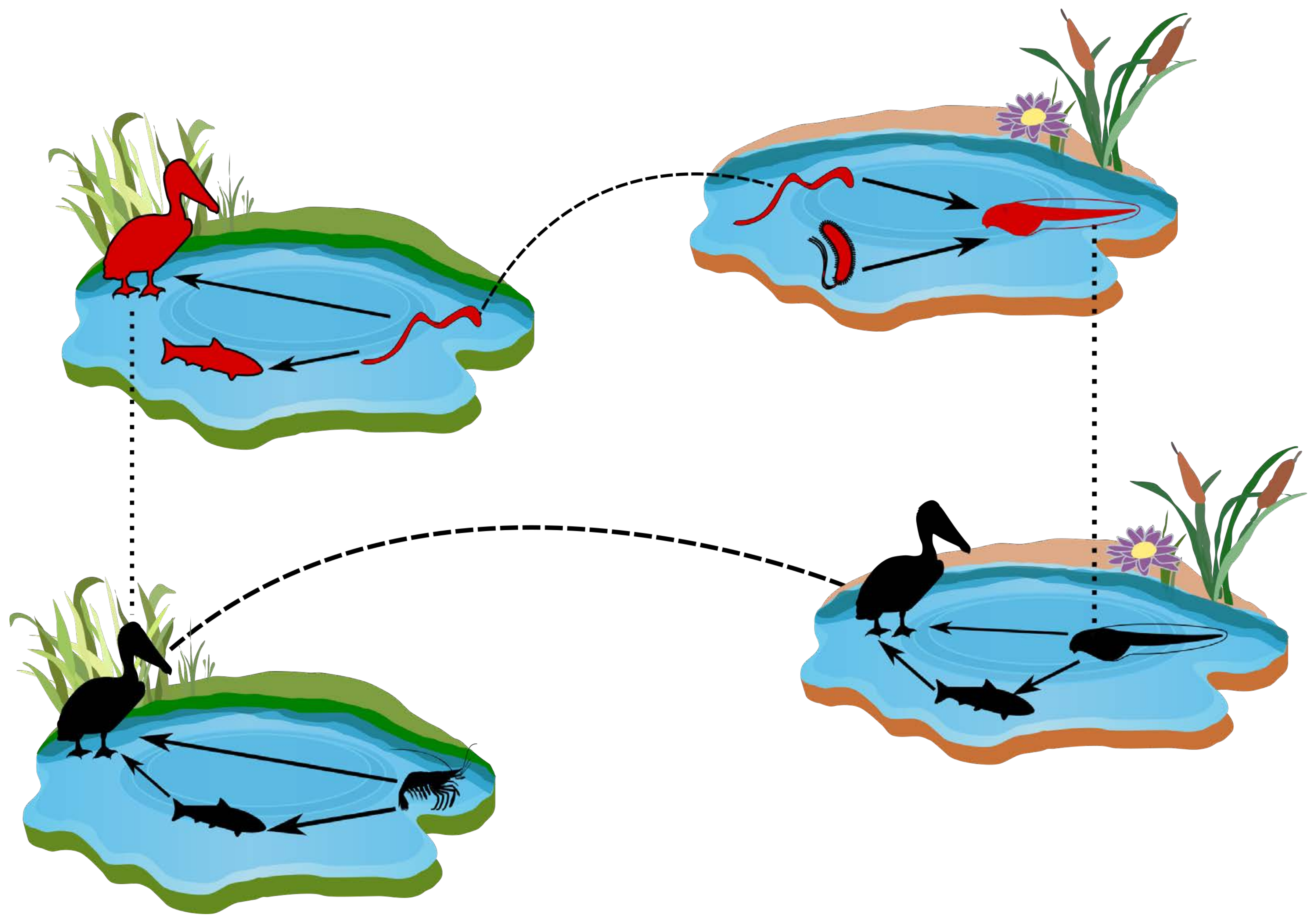


Frydman,…,Pilosof 2023 Methods Ecol Evol

Multipartite

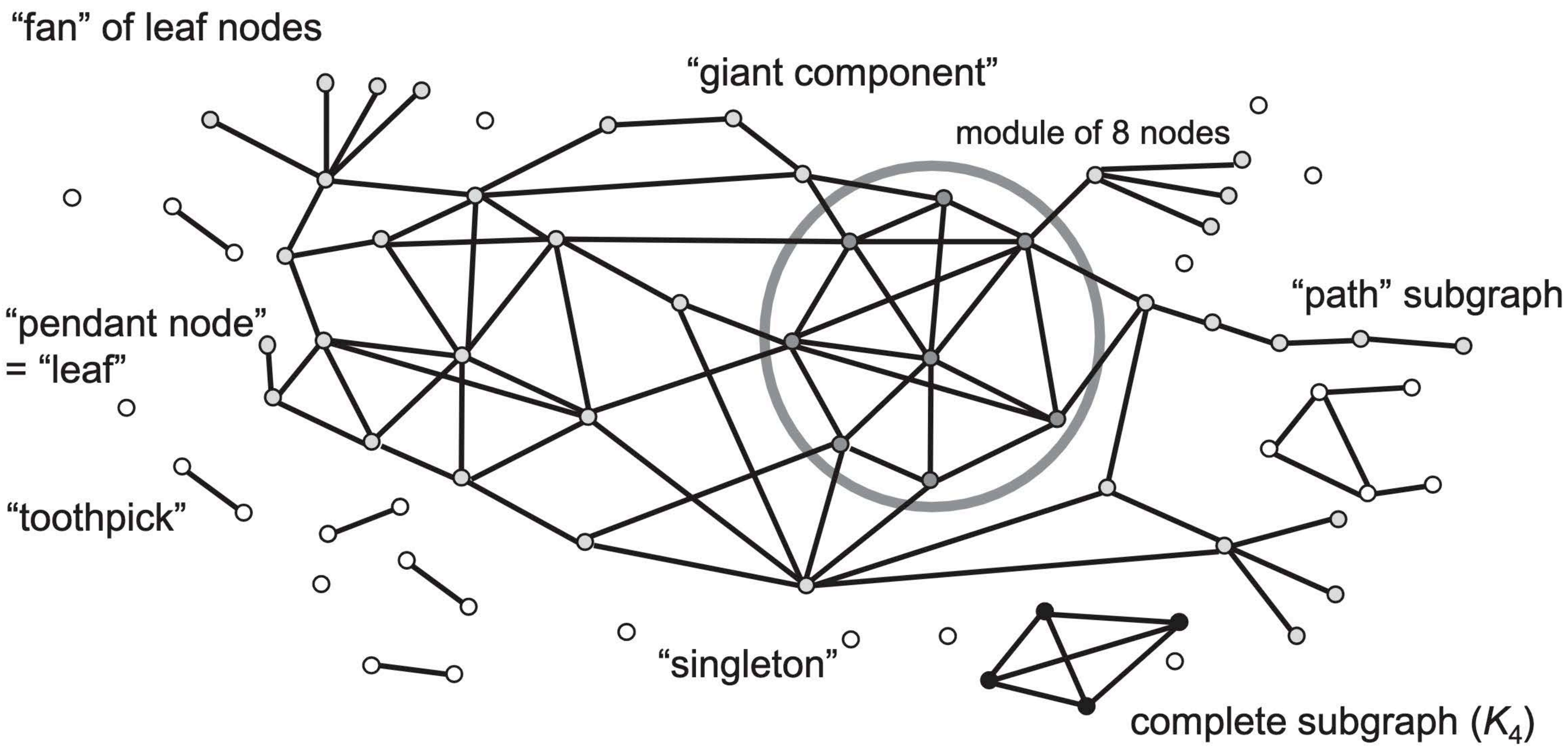
Multilayer





Pocock et al 2012 Science; Pilosof et al. 2017 Nature Ecol. Evol.

Quantifying structure



Node level

- Degree
- Centrality

Mesoscale level

- Motifs
- Modularity

Network level

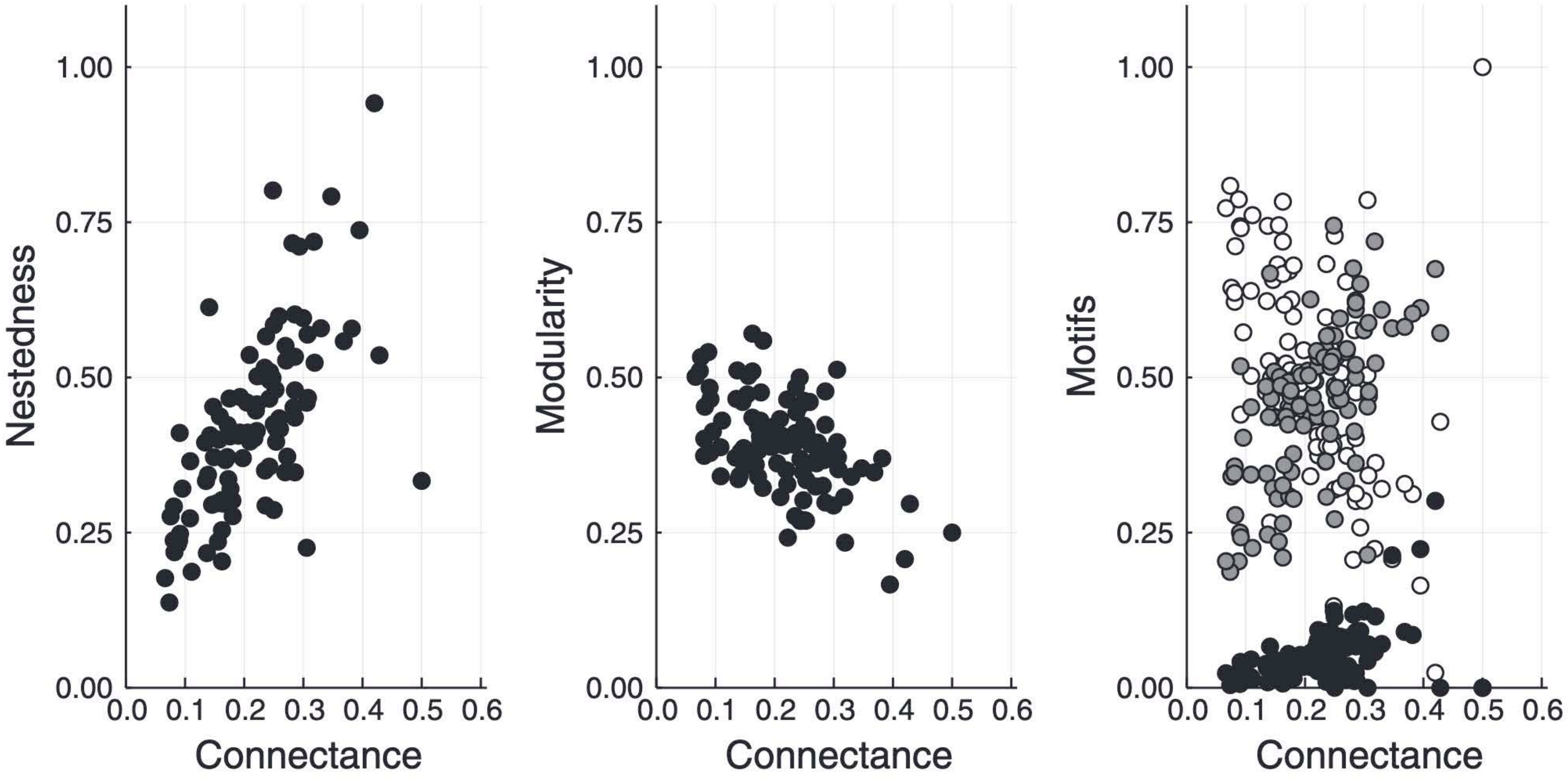
- Number of links
- Number of nodes
- Connectance
- Nestedness

Dale and Fortin 2021

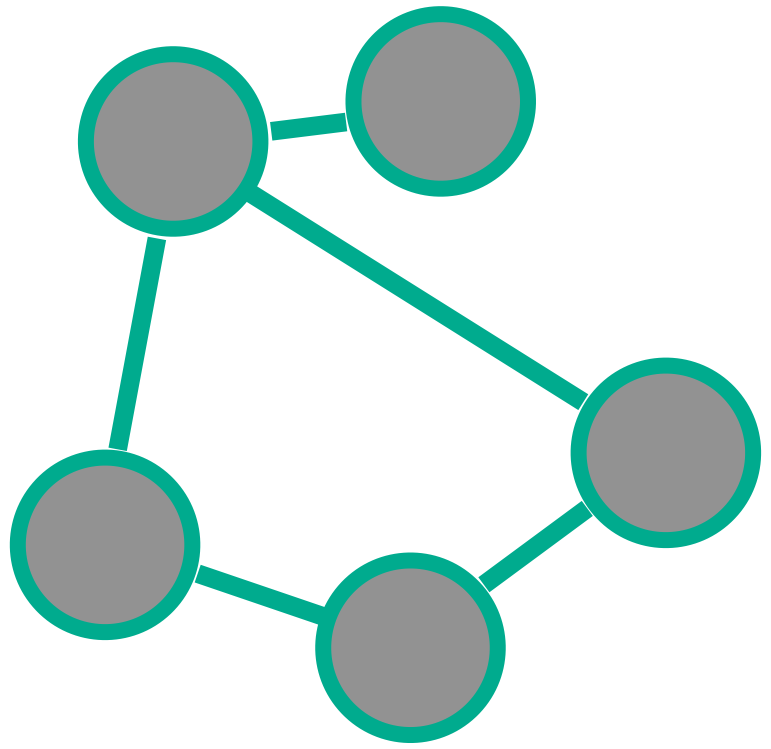
Connectance (density)

- L - # of links, S - matrix size (order)
- L/S - linkage density
- Connectance - number of links out of possible links, L/m. What is m?
  - Unipartite, undirected: $2L/S(S-1)$
  - Unipartite directed: $L/S(S-1)$
  - Bipartite: ?

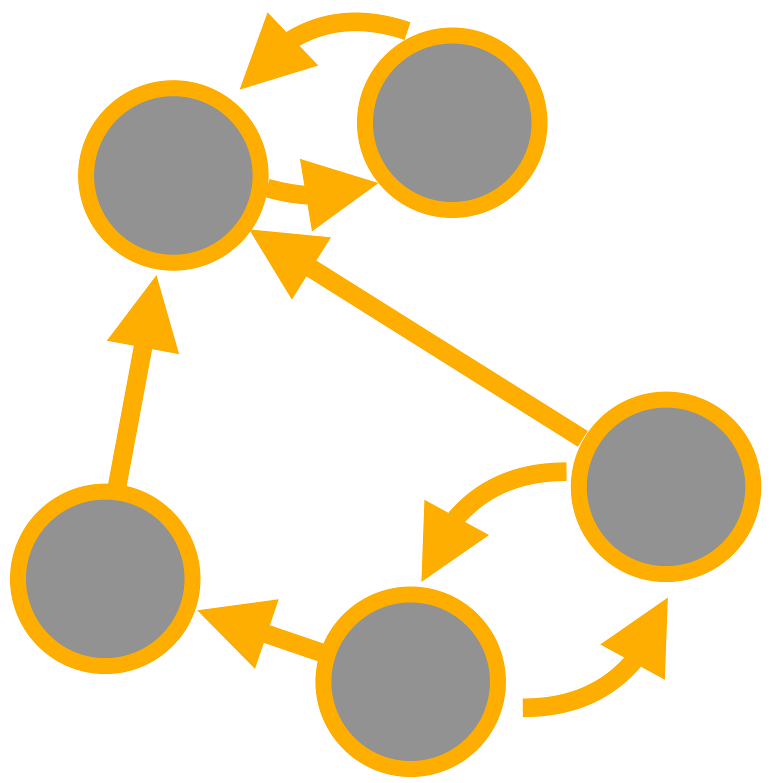
- First insights into the system.
- What affects L and m?
- Connectance correlates with other properties.



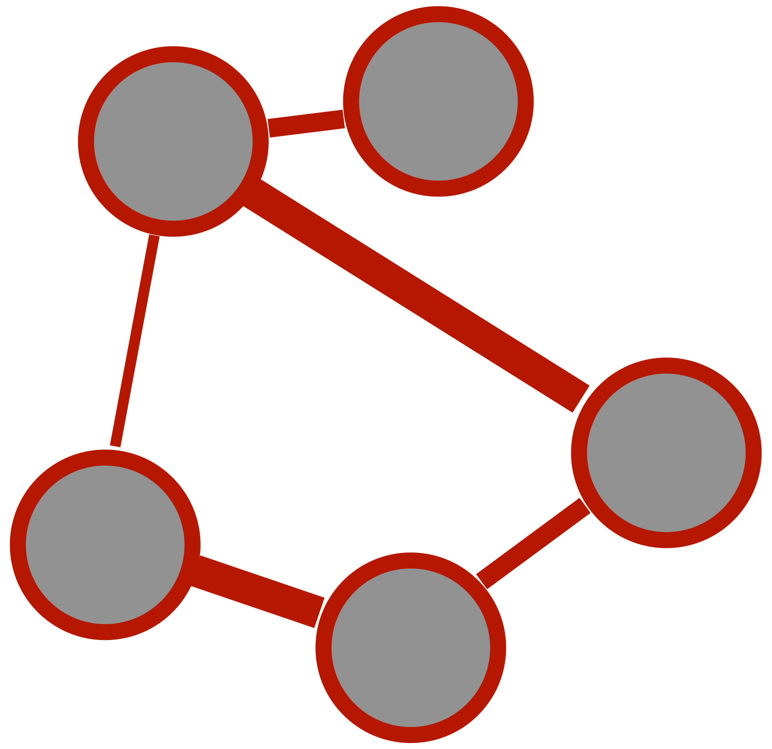
Fortuna et al 2010; Delmas et al 2019



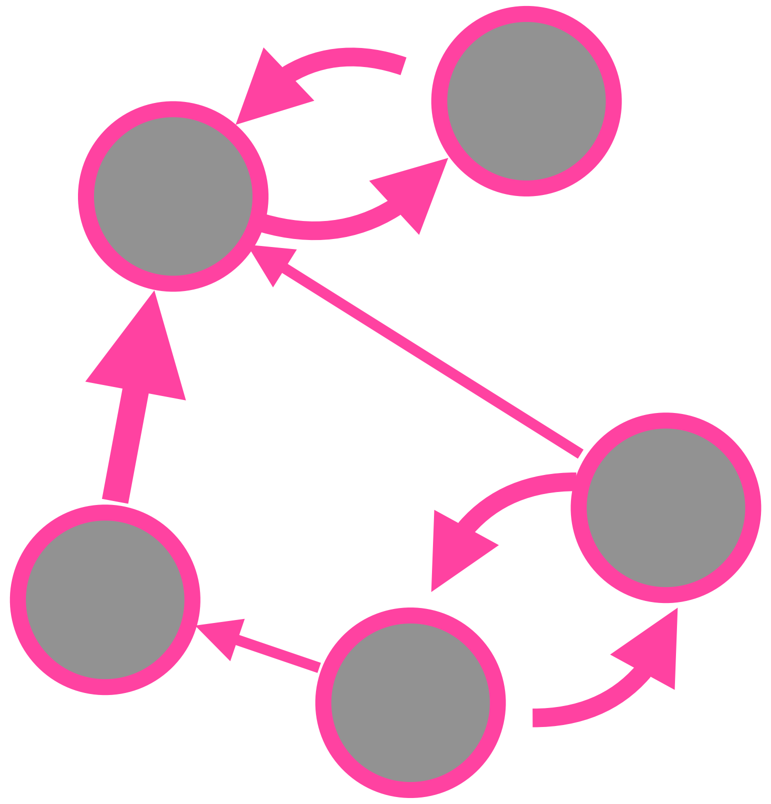
Binary (unweighted)
undirected



Binary (unweighted)
directed



Weighted
undirected



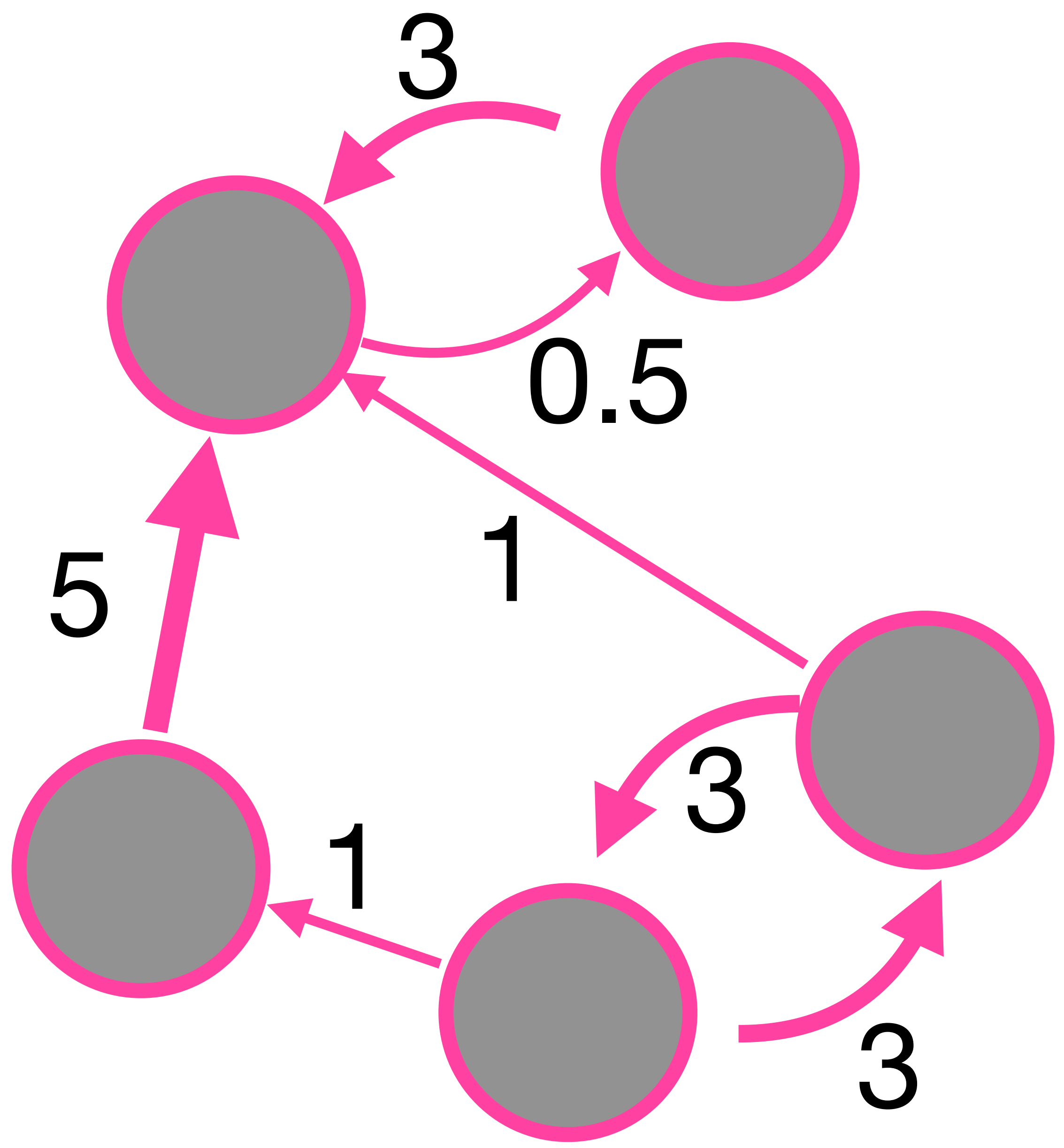
Weighted
directed

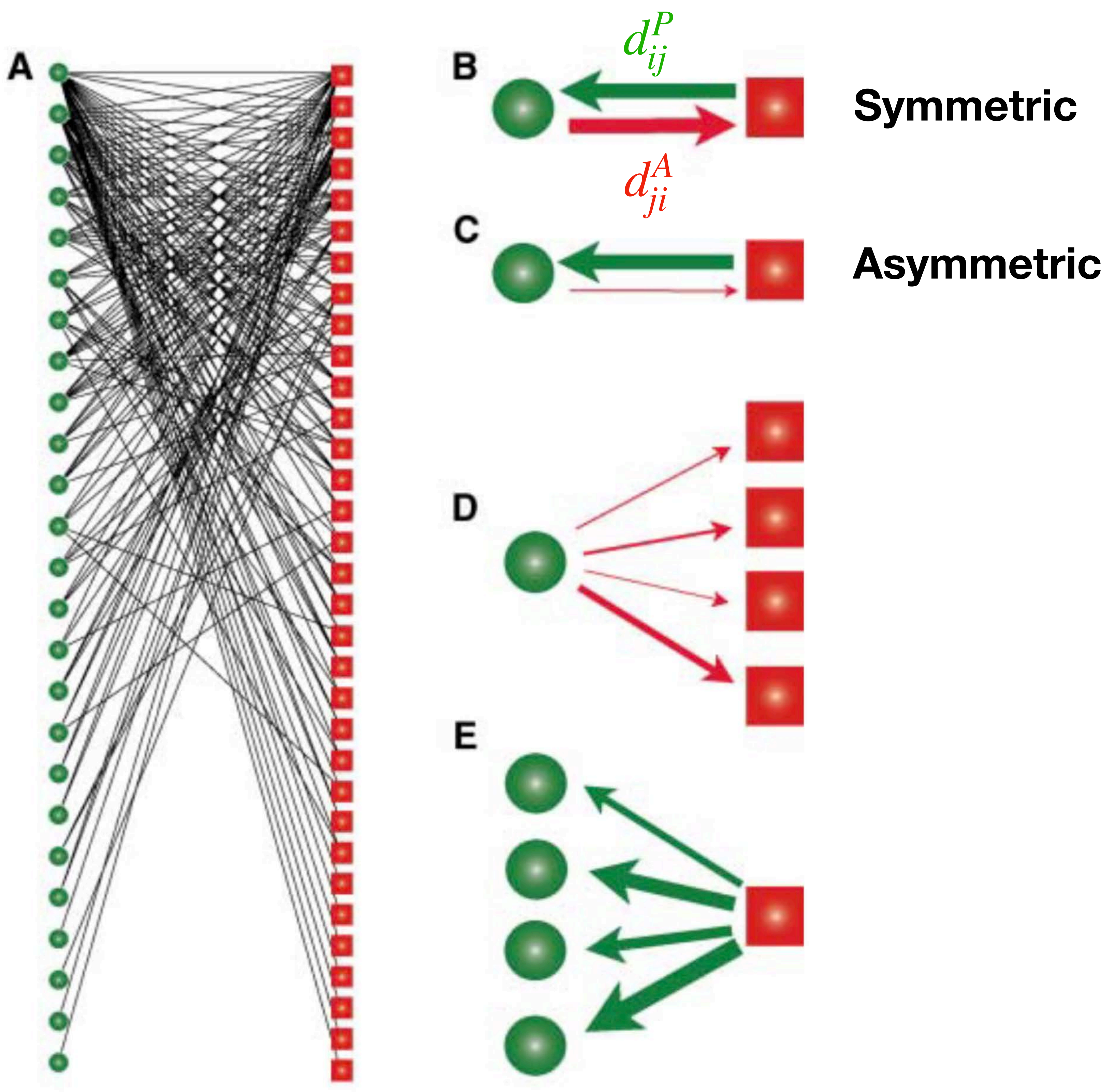
Degree: number of links a node has

In-degree: number of links pointing **to** a node

Out-degree: number of links pointing **from** a node

Strength: sum of link values of a node





Strength: sum of dependences

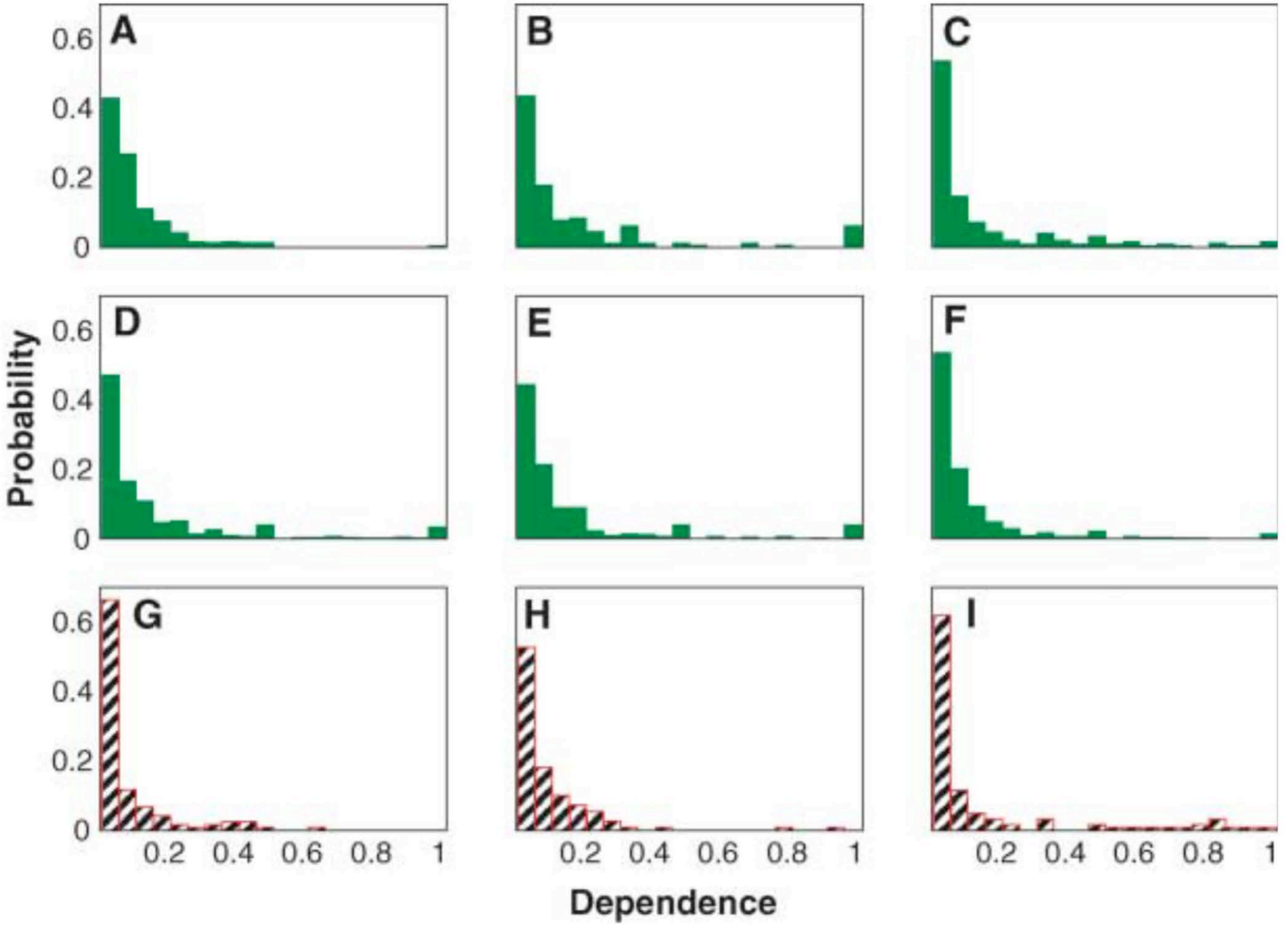


Fig. 2. Frequency distributions of dependence values within a mutualistic community. Green solid histograms (**A** to **F**) represent dependences of plants on pollinators, and red dashed histograms (**G** to **I**) represent dependences of seed dispersers on plants. See Database S1 for references and data sets.

$$AS_{ij} = \frac{|d^P_{ij} - d^A_{ji}|}{max(d^P_{ij}, d^A_{ji})}$$

d^P_{ij} — fraction of all visits by j going to this plant i.

d^A_{ji} — Fraction of all animal visits to plant i coming from j

Bascompte. Jordano, Olesen 2006, Science

Measuring specialization

$$RR = \frac{R - r}{R - 1}$$: number of links established out of possible ones

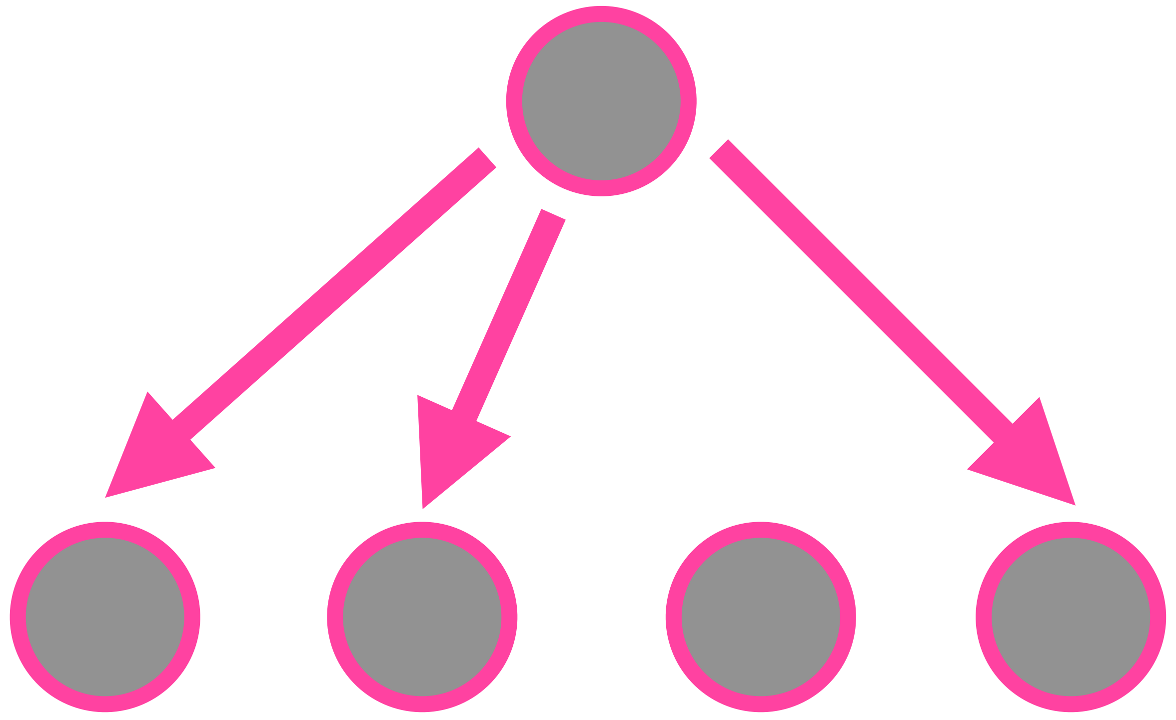
$$\mathrm{PDI} = \frac{\sum_{i=2}^{R} (\mathbf{P}_1 - \mathbf{P}_i)}{R - 1}$$

Contrast the strongest link (P1) with remaining resources (Pi)

$$\mathrm{HS} = \frac{\sum_{i=1}^{R} \left[\frac{\mathbf{P}_i}{\mathbf{T}} \ln\left(\frac{\mathbf{P}_i}{\mathbf{T}} \right) \right]}{\ln(R)} + 1.$$

Not defined for Pi=0

These (and other) measures of specificity capture different aspects of interactions. Read further to select the most appropriate one.



$$P = (s1, s2, s3, \ldots, R)$$

r: number of non-zero elements

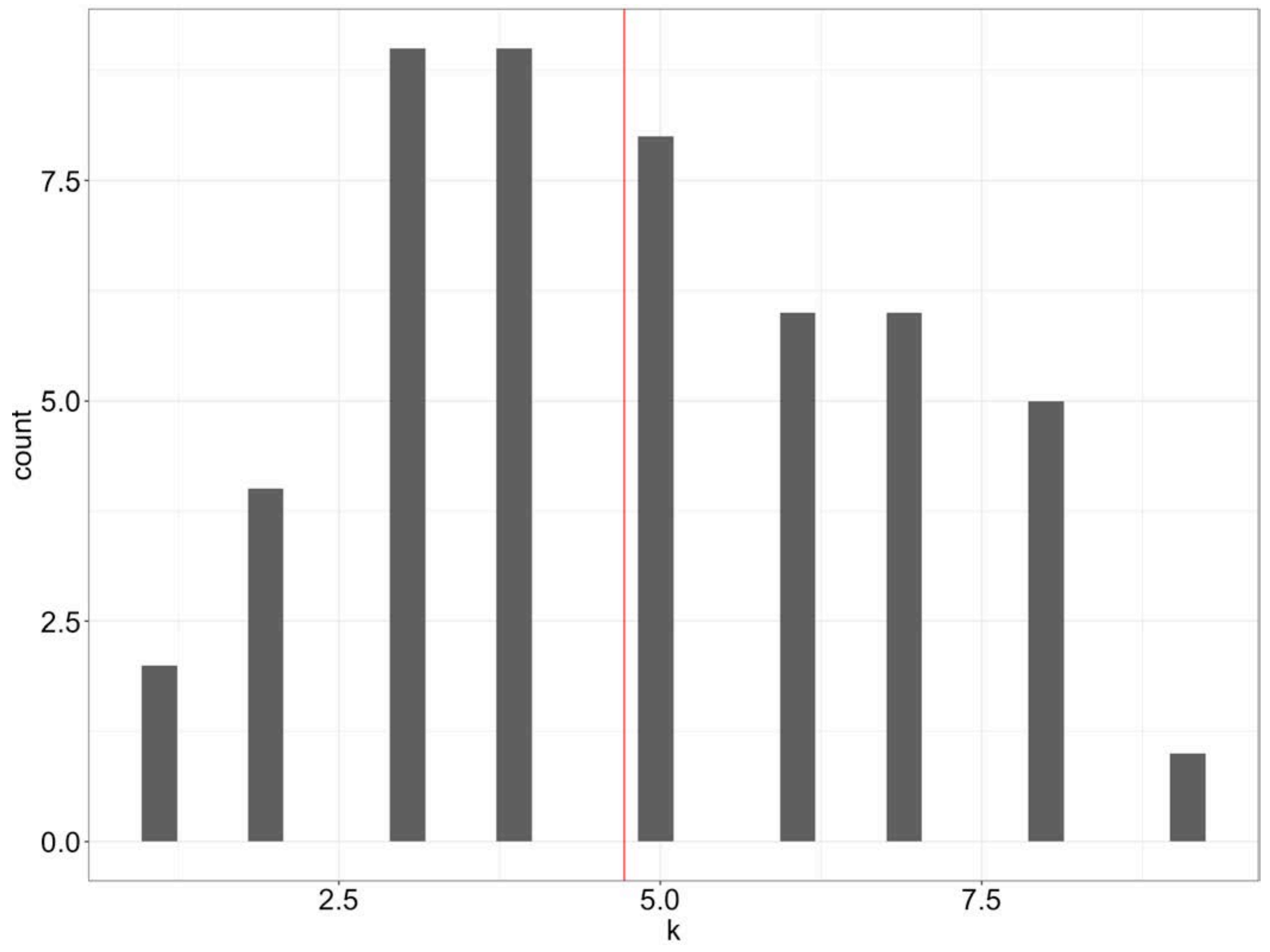
$$T = \sum_{i=1}^{R} P$$

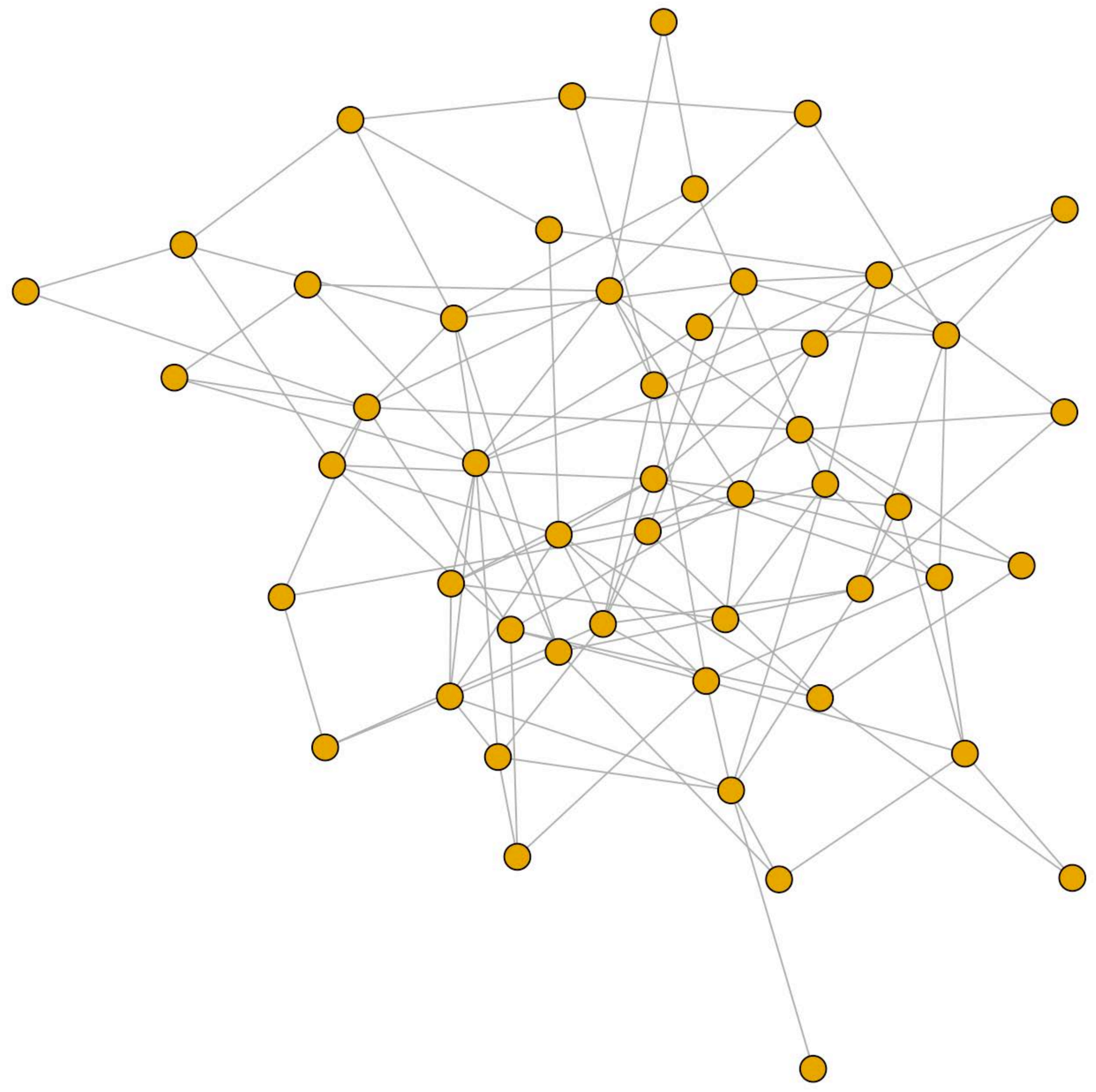
Normalization such that
0: absoulte generality; 1: absolute specificity

Poisot et al 2012, Methods Ecol Evol

Degree distributions: ER

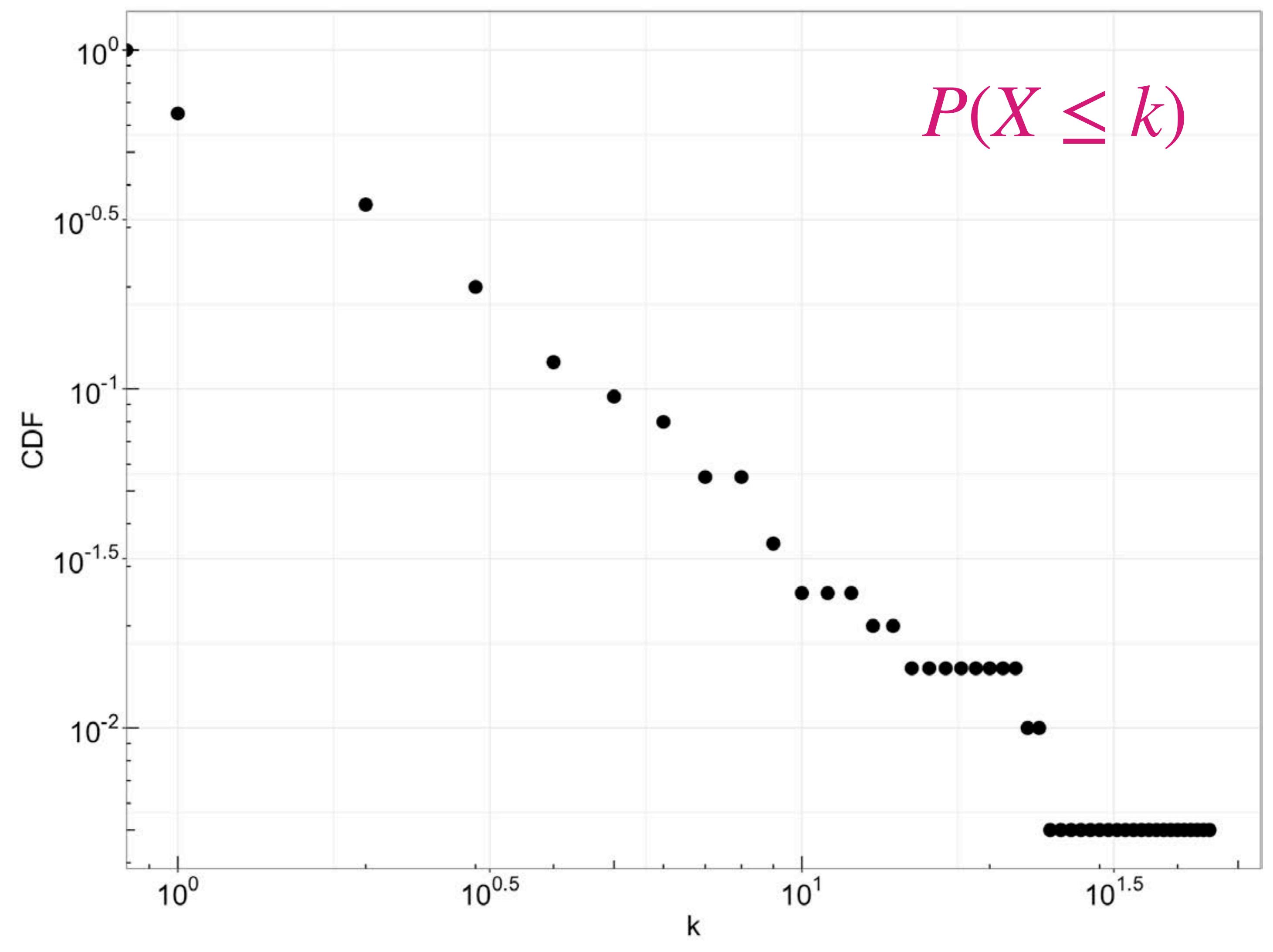
- $p(k)$: Proportion of nodes with degree k.
- Erdos-Reyni graph: edges are added randomly with probability m .
  - Binomial degree distribution (Poisson for large n).
  - How many parameters are needed to describe the graph?

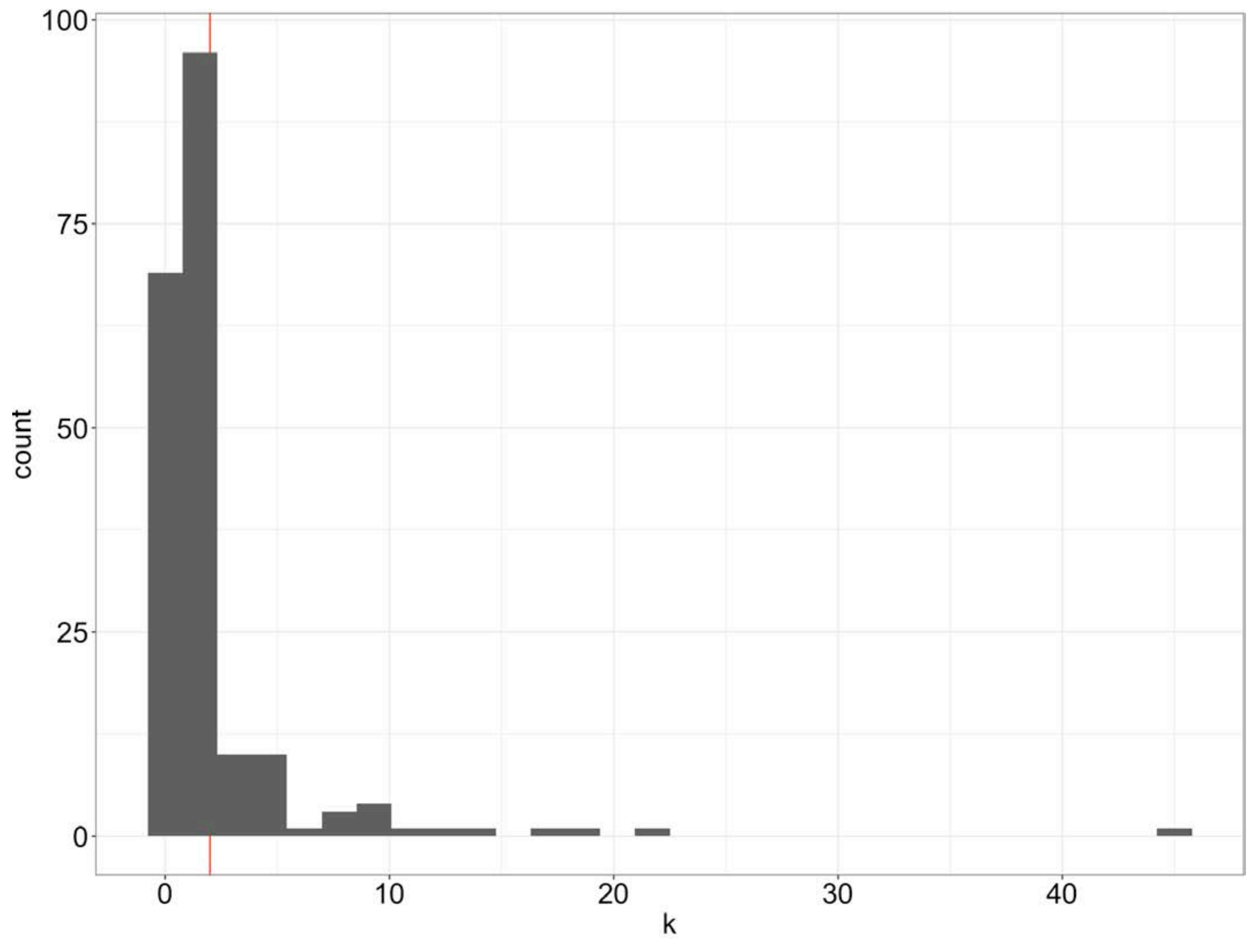


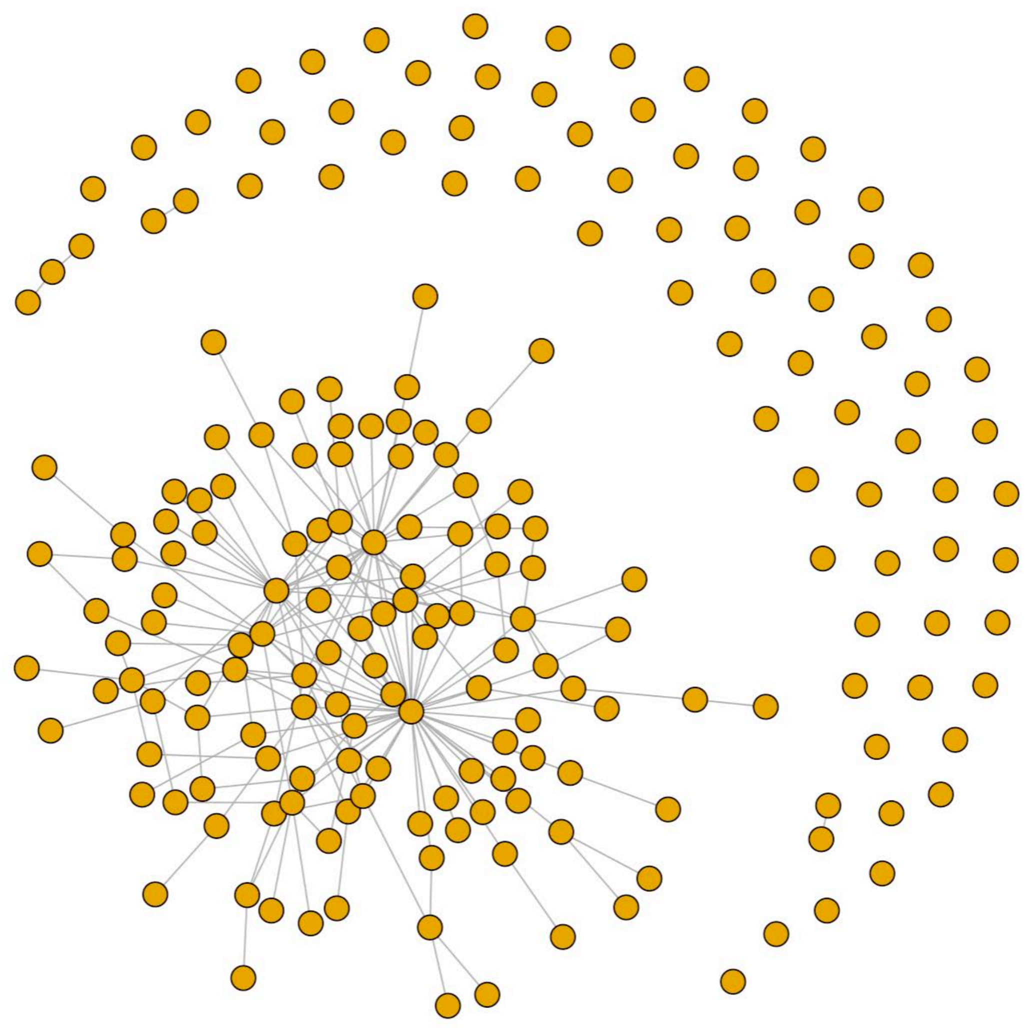


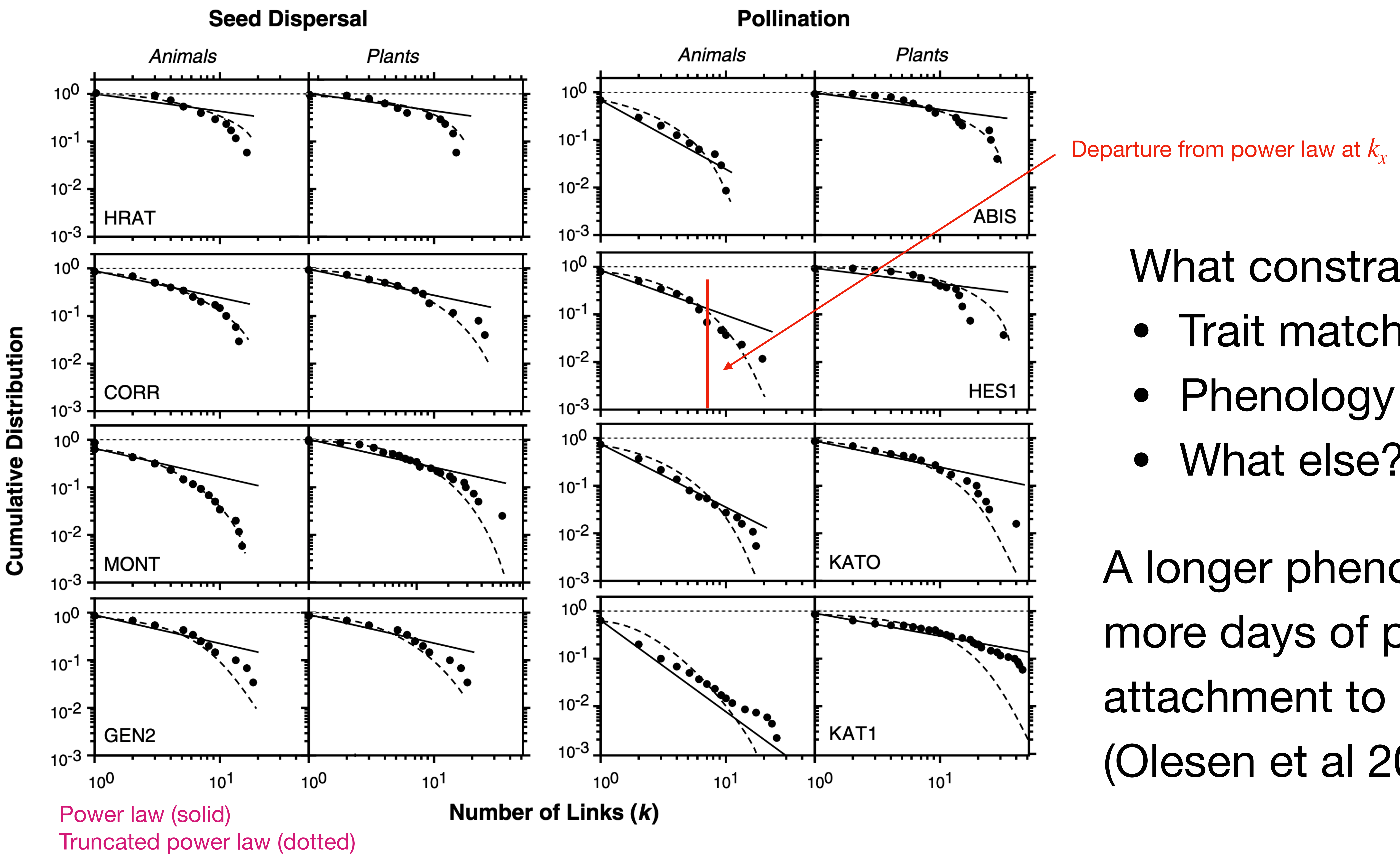
Degree distributions: power-law graphs

- Power-law graph (or a scale-free network): $p(k) \propto Ck^{-\gamma}$, where $p(k)$ is the probability of observing a node with degree k.
- rich gets richer / preferential attachment (Barabási-Albert 1999)
  - Scaling k does not change the relationship.
  - Hubs are typical
  - What is the exponent
  - What does deviation from the power-law tells us?
  - What are the dynamical implications: robustness, immunization.









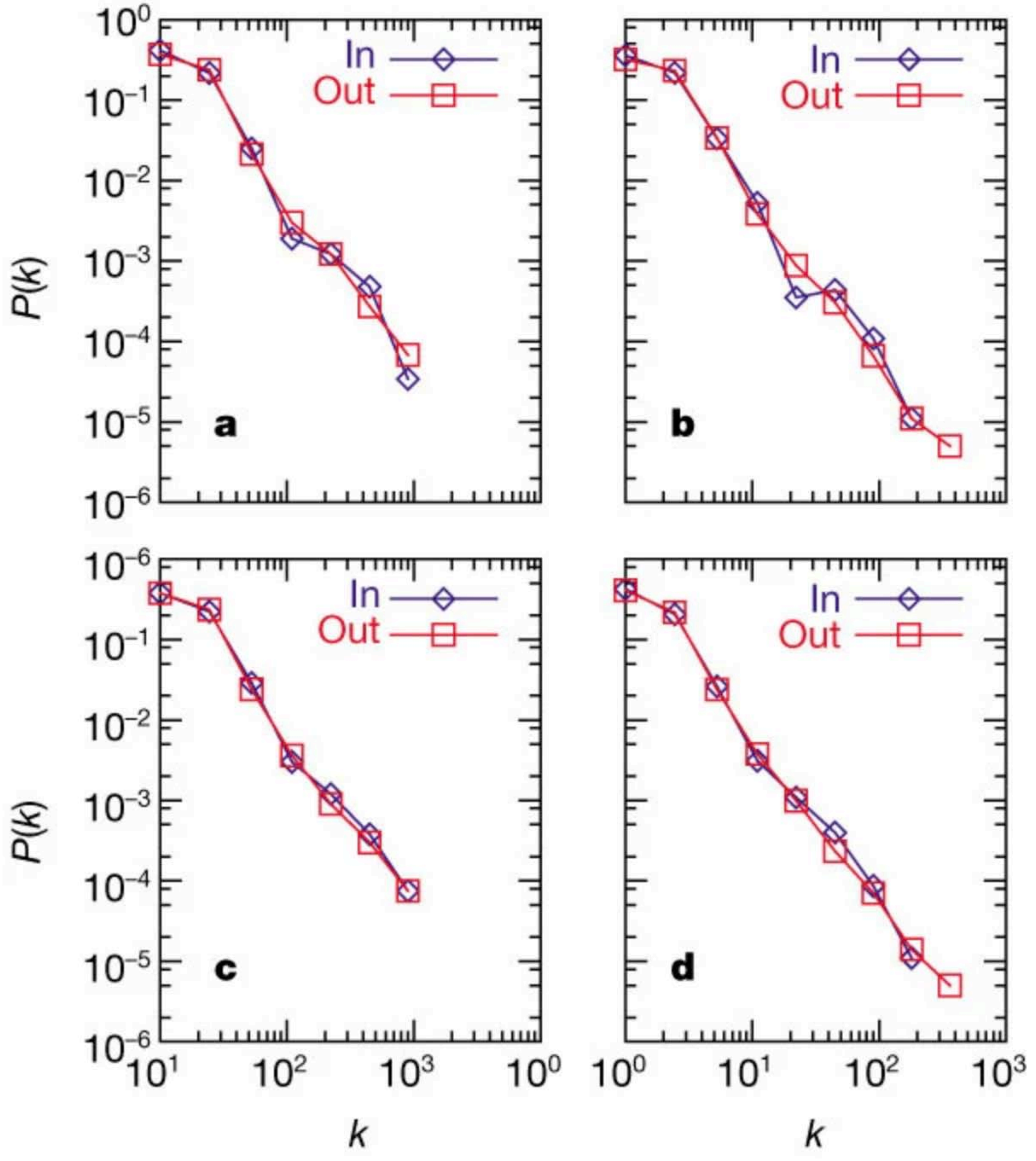
What constraints links?

- Trait matching
- Phenology
- What else?

A longer phenophase leads to more days of potential attachment to new species. (Olesen et al 2008)

Metabolic networks are scale free

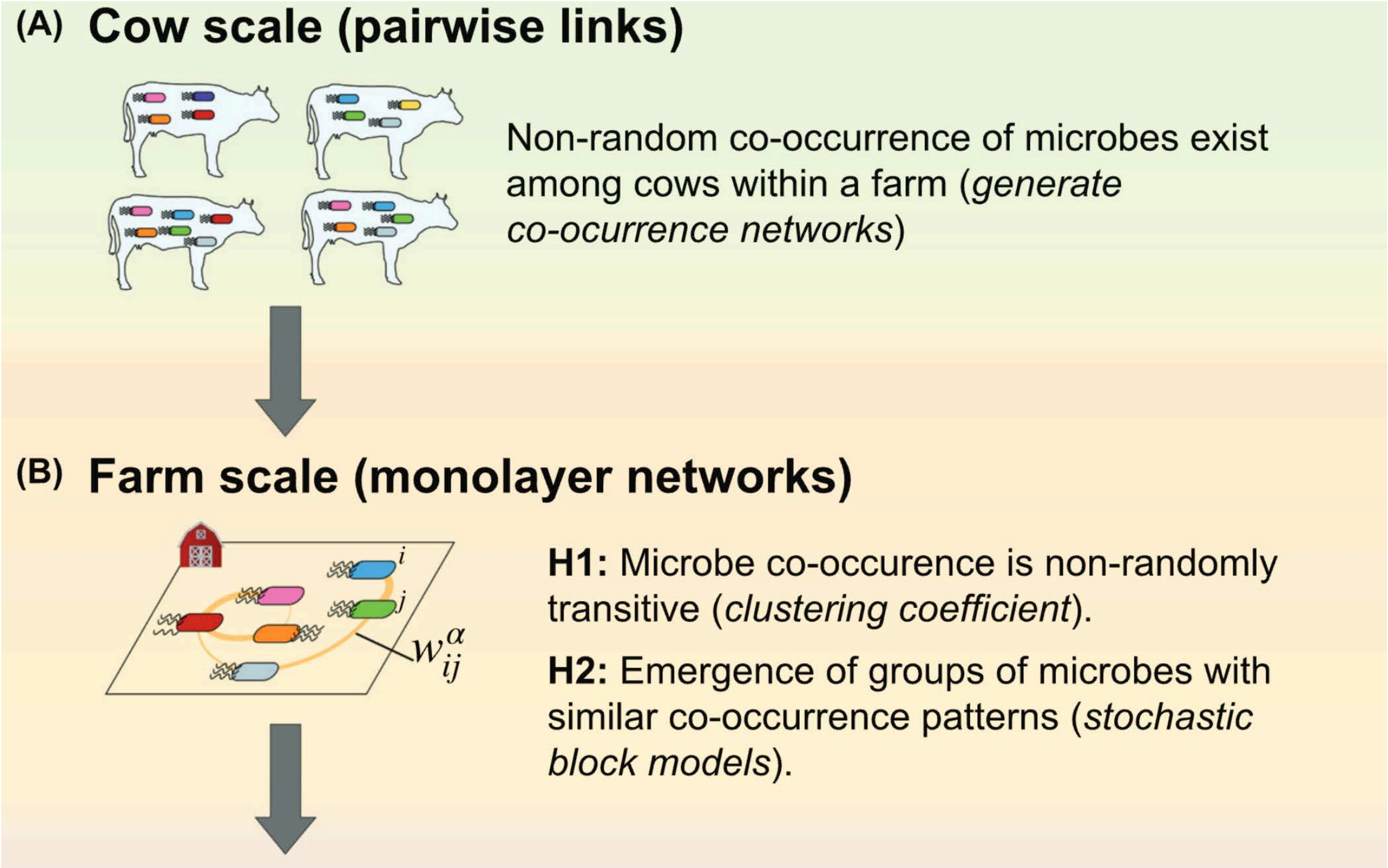
Figure 2: Connectivity distributions *P*(*k*) for substrates.

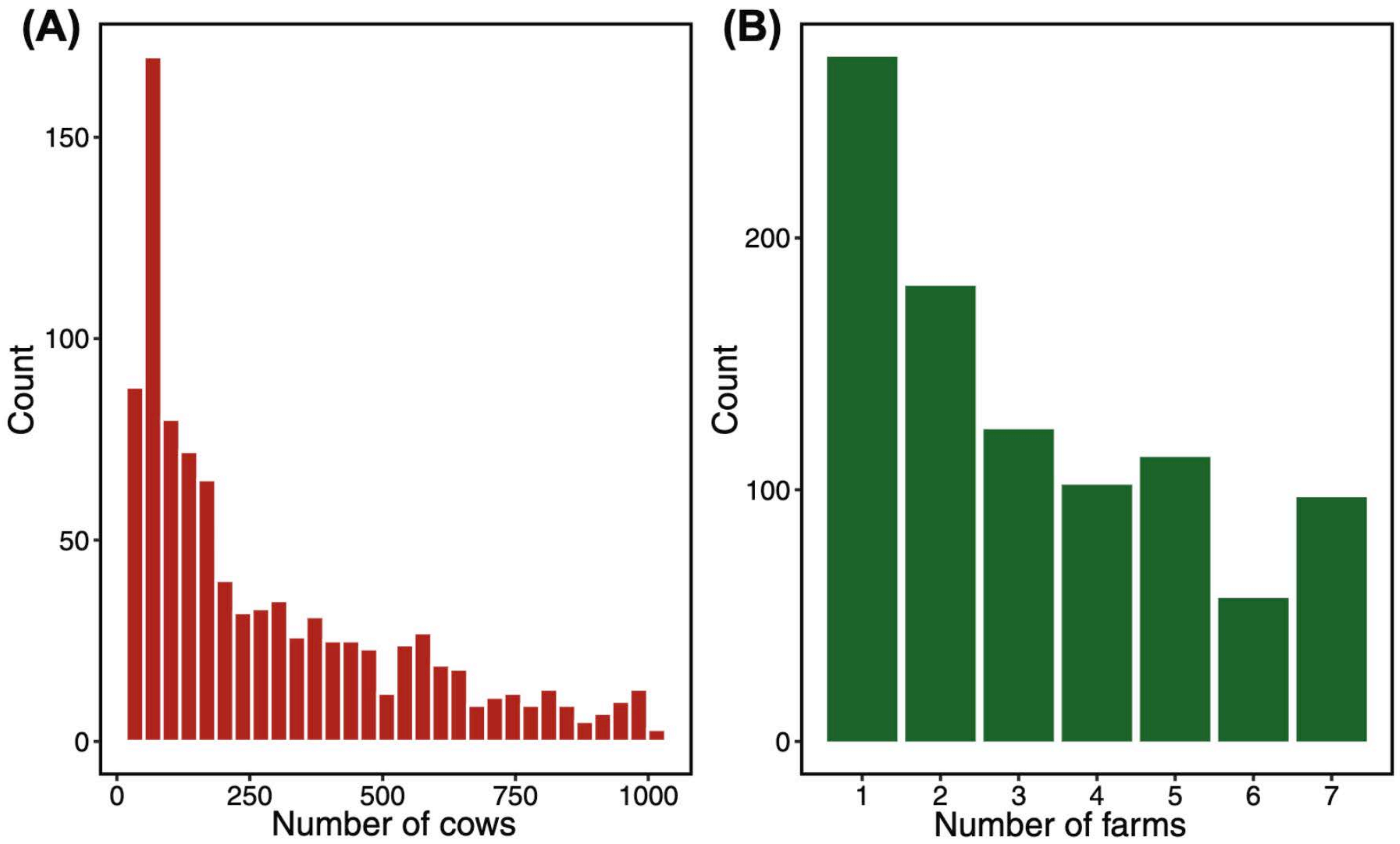


a, *Archaeoglobus fulgidus* (archae); **b**, *E. coli* (bacterium); **c**, *Caenorhabditis elegans* (eukaryote), shown on a log–log plot, counting separately the incoming (In) and outgoing links (Out) for each substrate. k_{in} (k_{out}) corresponds to the number of reactions in which a substrate participates as a product (educt). The characteristics of the three organisms shown in **a**–**c** and the exponents γ_{in} anc γ_{out} for all organisms are given in Table 1 of the Supplementary Information. **d**, The connectivity distribution averaged over all 43 organisms.

Jeong et al 2000, Nature

Microbe distribution in hosts is skewed

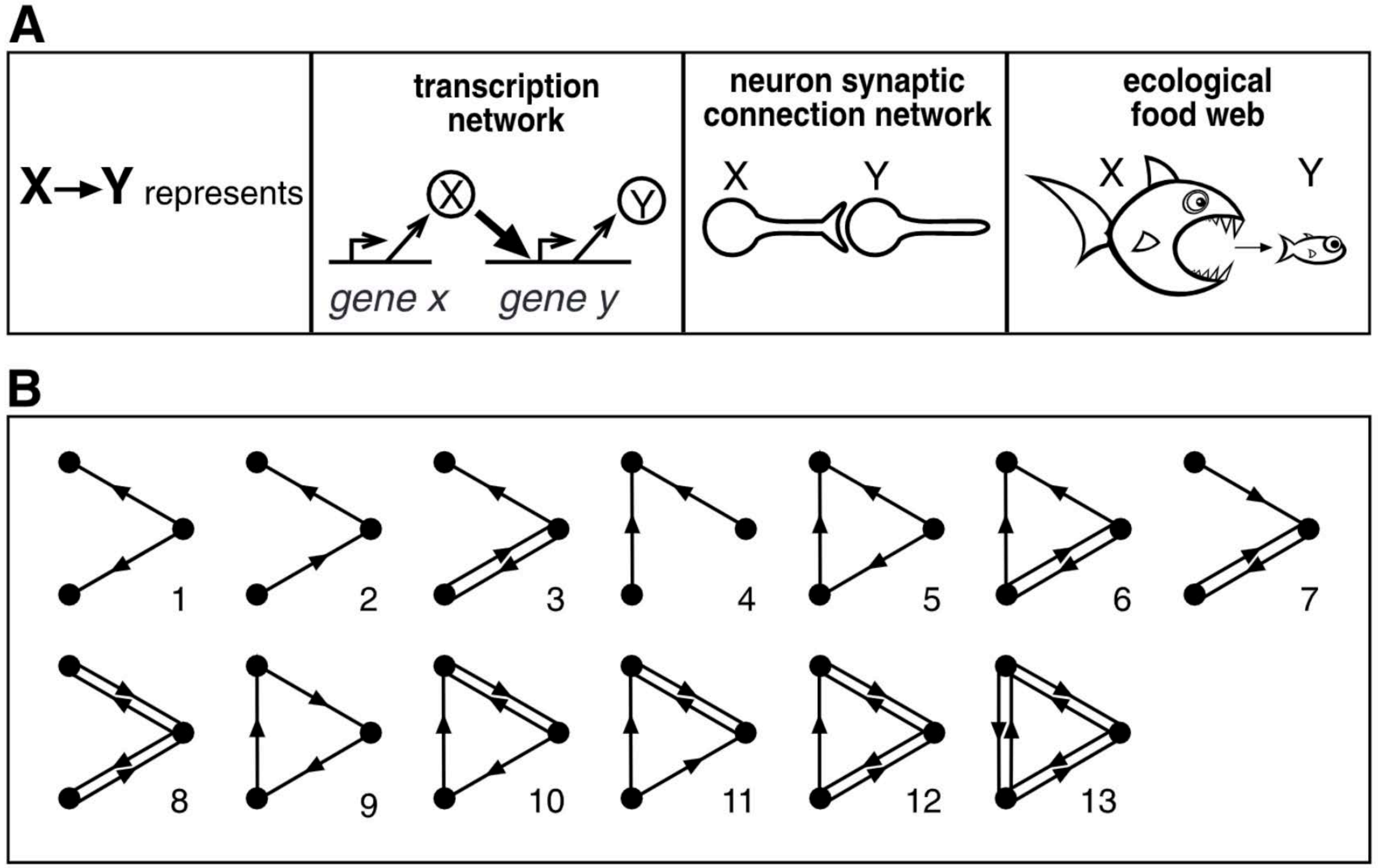


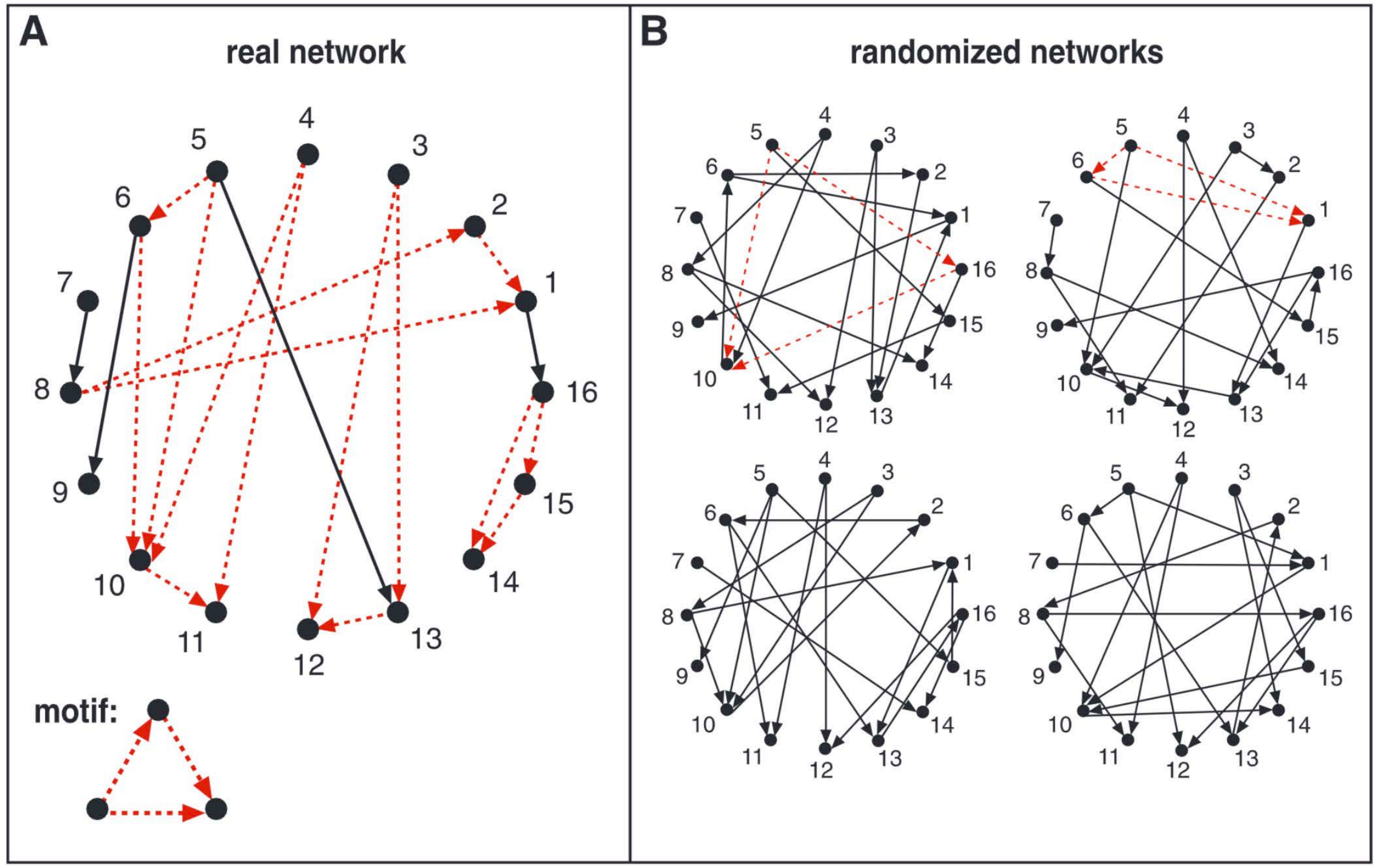


Galai et al 2024

Motifs

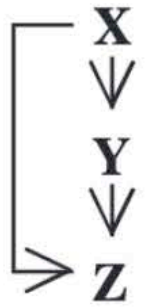
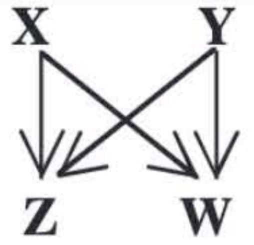
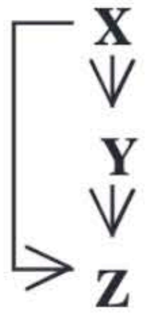
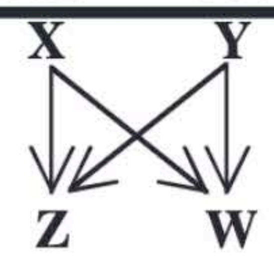
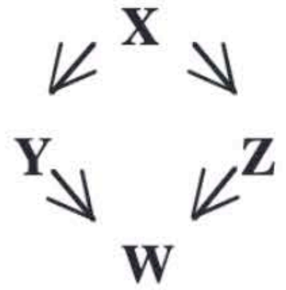
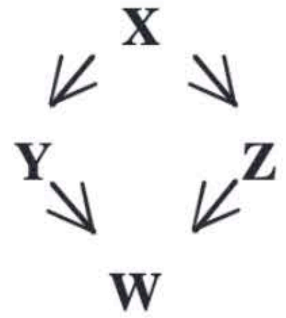
- n-node subgraphs
- Motifs and motif profiles Can be indicative of generative processes and function.





Milo et al 2002 Science

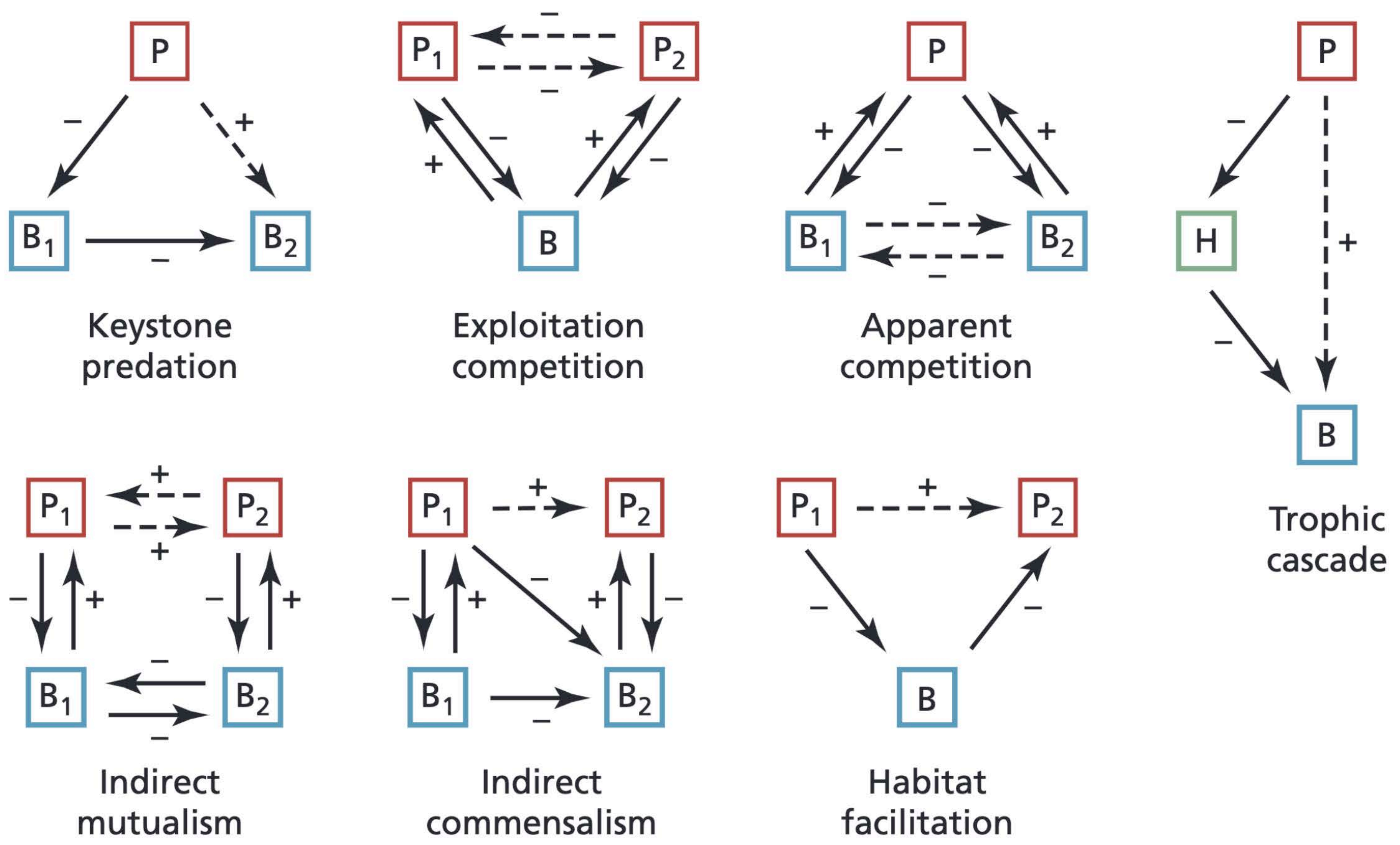
Motifs define broad classes of networks?

Network	Nodes	Edges	N_{real}	$N_{rand} \pm$ SD	*Z* score	N_{real}	$N_{rand} \pm$ SD	Z score	N_{real}	$N_{rand} \pm$ SD	*Z* score
Gene regulation (transcription)					**Feed-forward loop**			**Bi-fan**			
E. coli	424	519	40	7 ± 3	10	203	47 ± 12	13			
*S. cerevisiae**	685	1,052	70	11 ± 4	14	1812	300 ± 40	41			
Neurons					**Feed-forward loop**			**Bi-fan**			**Bi-parallel**
C. elegans†	252	509	125	90 ± 10	3.7	127	55 ± 13	5.3	227	35 ± 10	20
Food webs					**Three chain**			**Bi-parallel**			
Little Rock	92	984	3219	3120 ± 50	2.1	7295	2220 ± 210	25			
Ythan	83	391	1182	1020 ± 20	7.2	1357	230 ± 50	23			
St. Martin	42	205	469	450 ± 10	NS	382	130 ± 20	12			
Chesapeake	31	67	80	82 ± 4	NS	26	5 ± 2	8			
Coachella	29	243	279	235 ± 12	3.6	181	80 ± 20	5			
Skipwith	25	189	184	150 ± 7	5.5	397	80 ± 25	13			
B. Brook	25	104	181	130 ± 7	7.4	267	30 ± 7	32			

Milo et al 2002 Science

Triadic “modules” are the building blocks of ecological communities

Motifs provide a smooth transition into community ecology theory

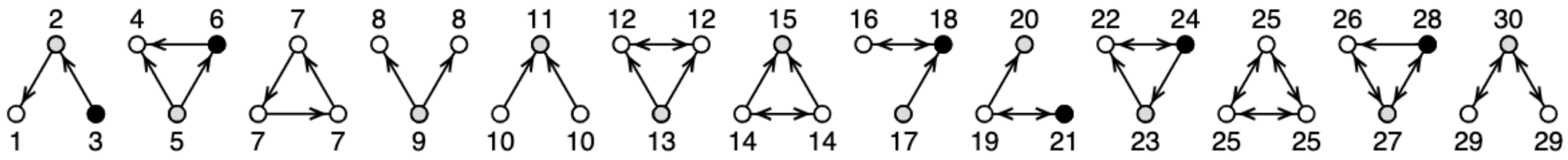


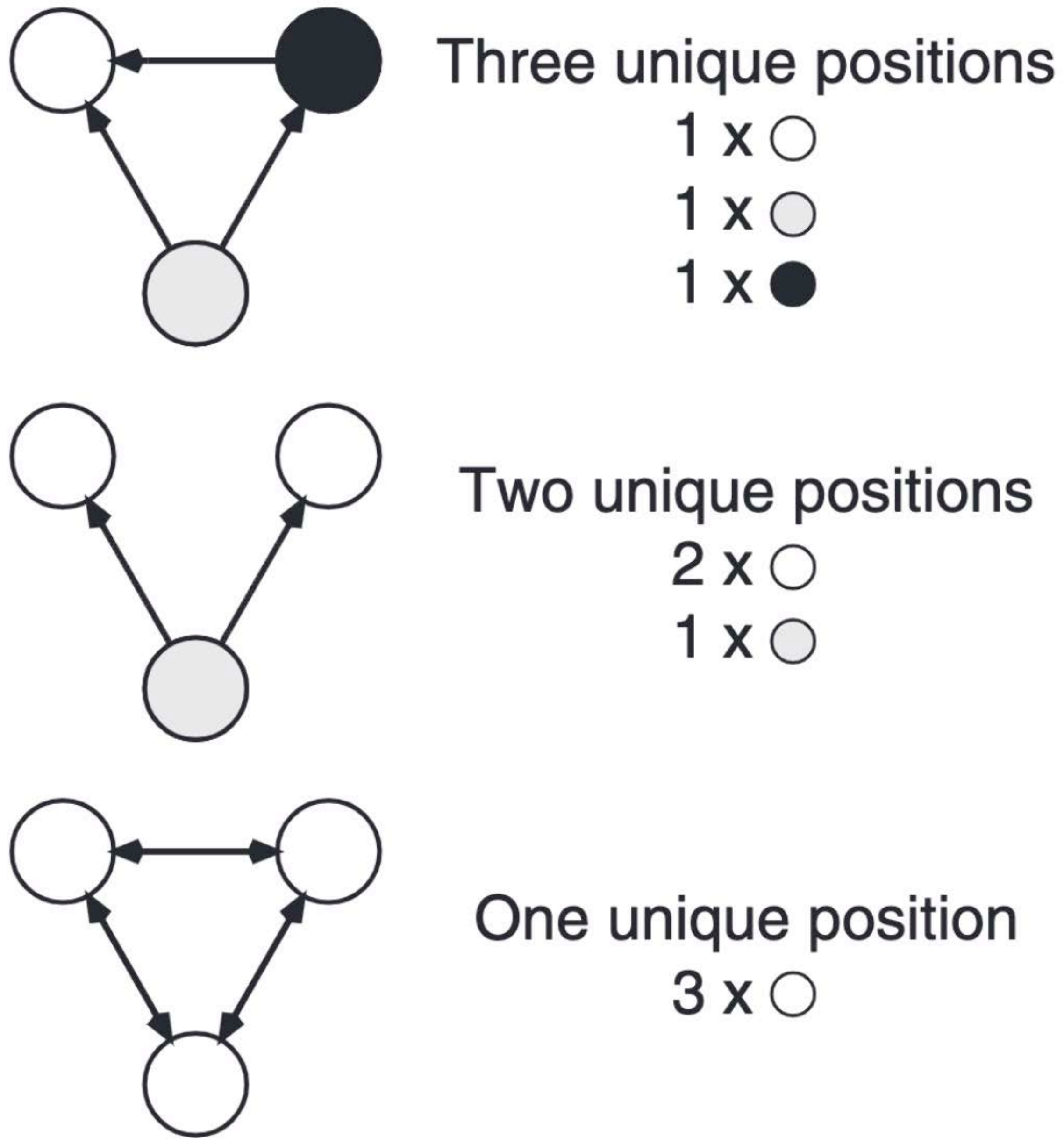
Mittelbach 2019 Community Ecology

Species roles can be defined as the position in a motif (30 unique positions).

Frequency that species I appears in position j.

Species-specific motif profiles.





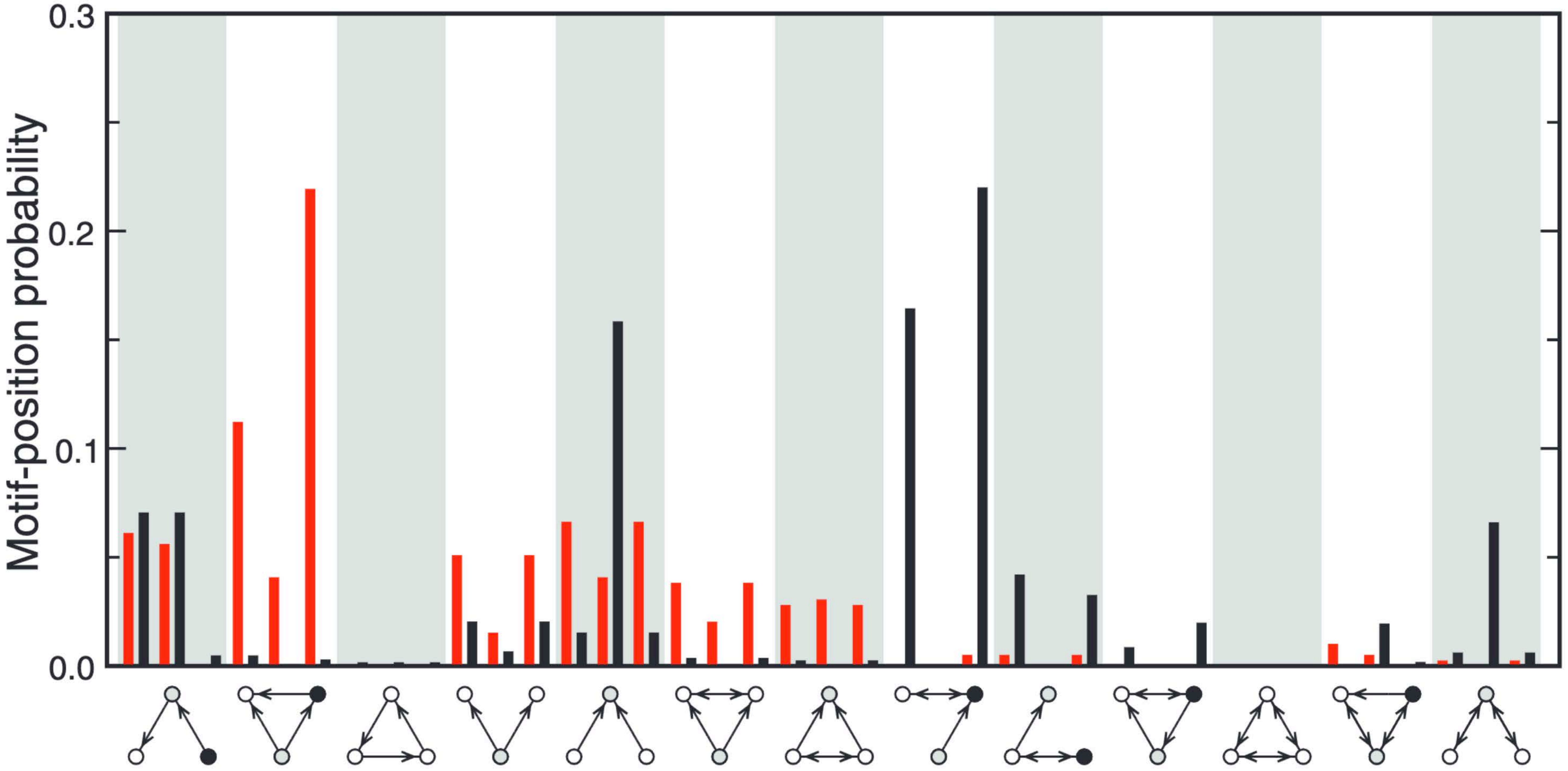
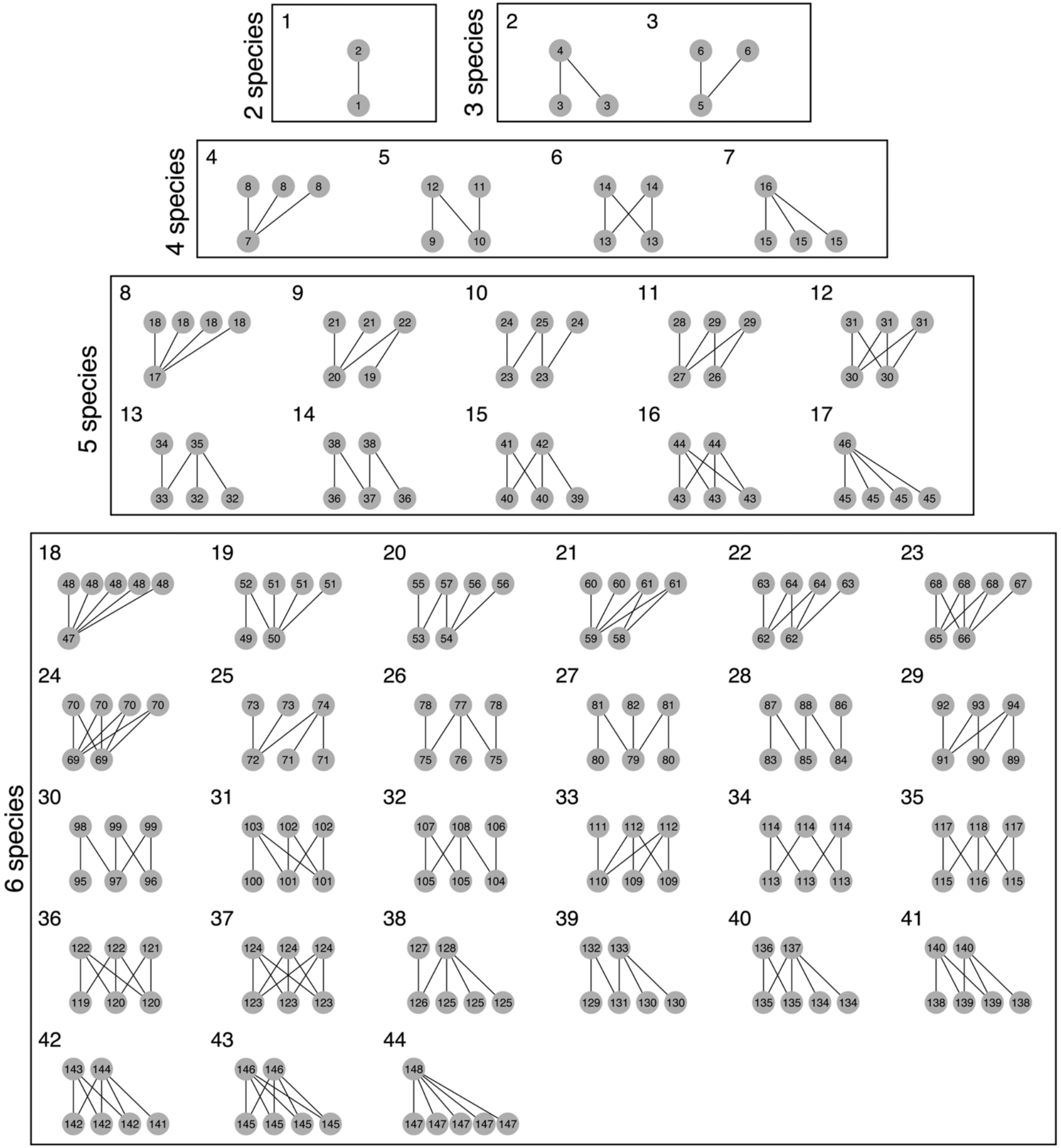


Fig. 2. Species differ in their tendency to appear in distinct motif positions. We show the species-specific motif profiles $\vec{f_i}$ for two different species from the empirical webs (red and black bars, respectively). The height of each bar is equal to the probability f_{ij} that the species appears in the position found immediately below.

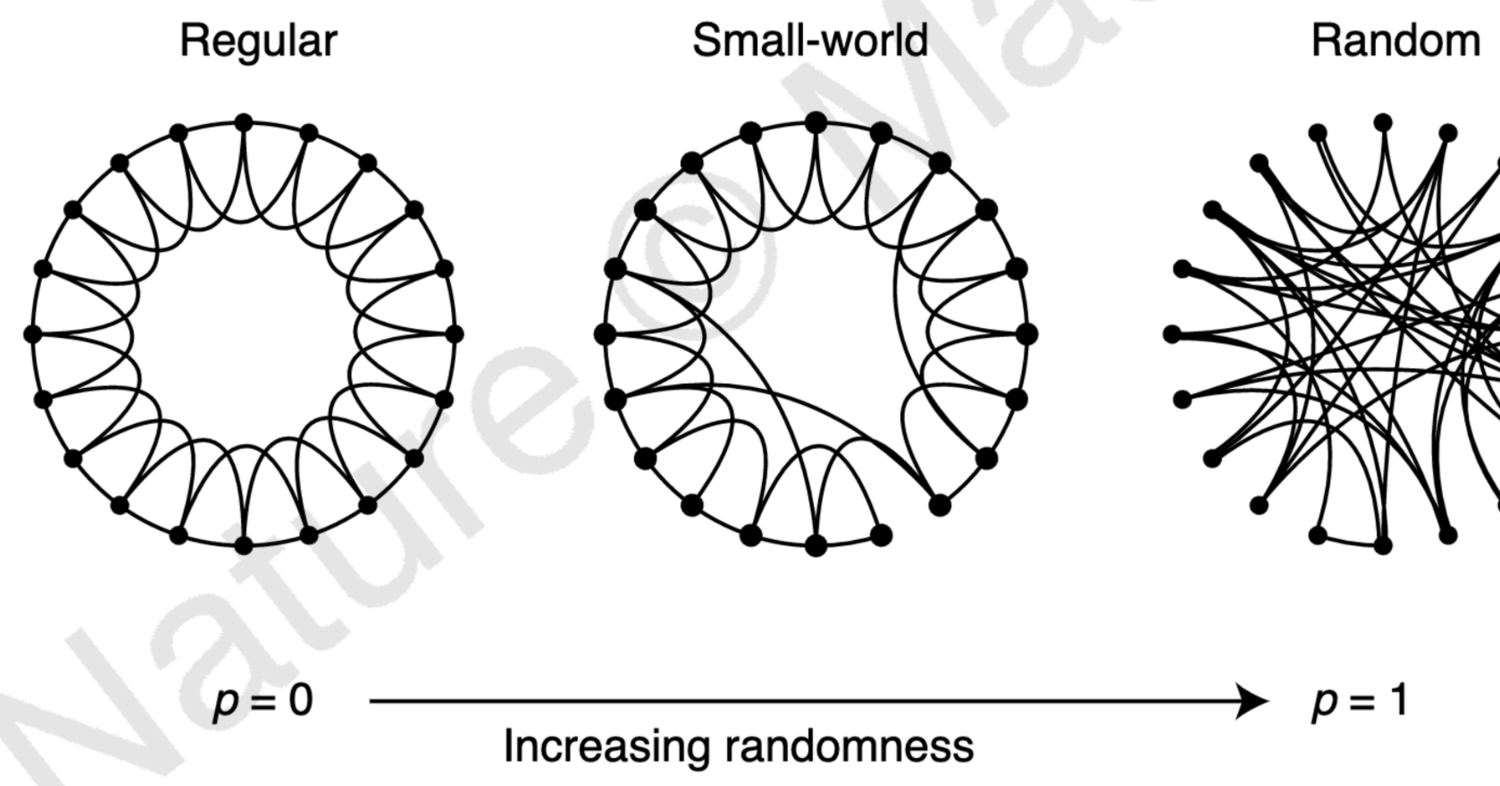
Stouffer et al 2012, Science

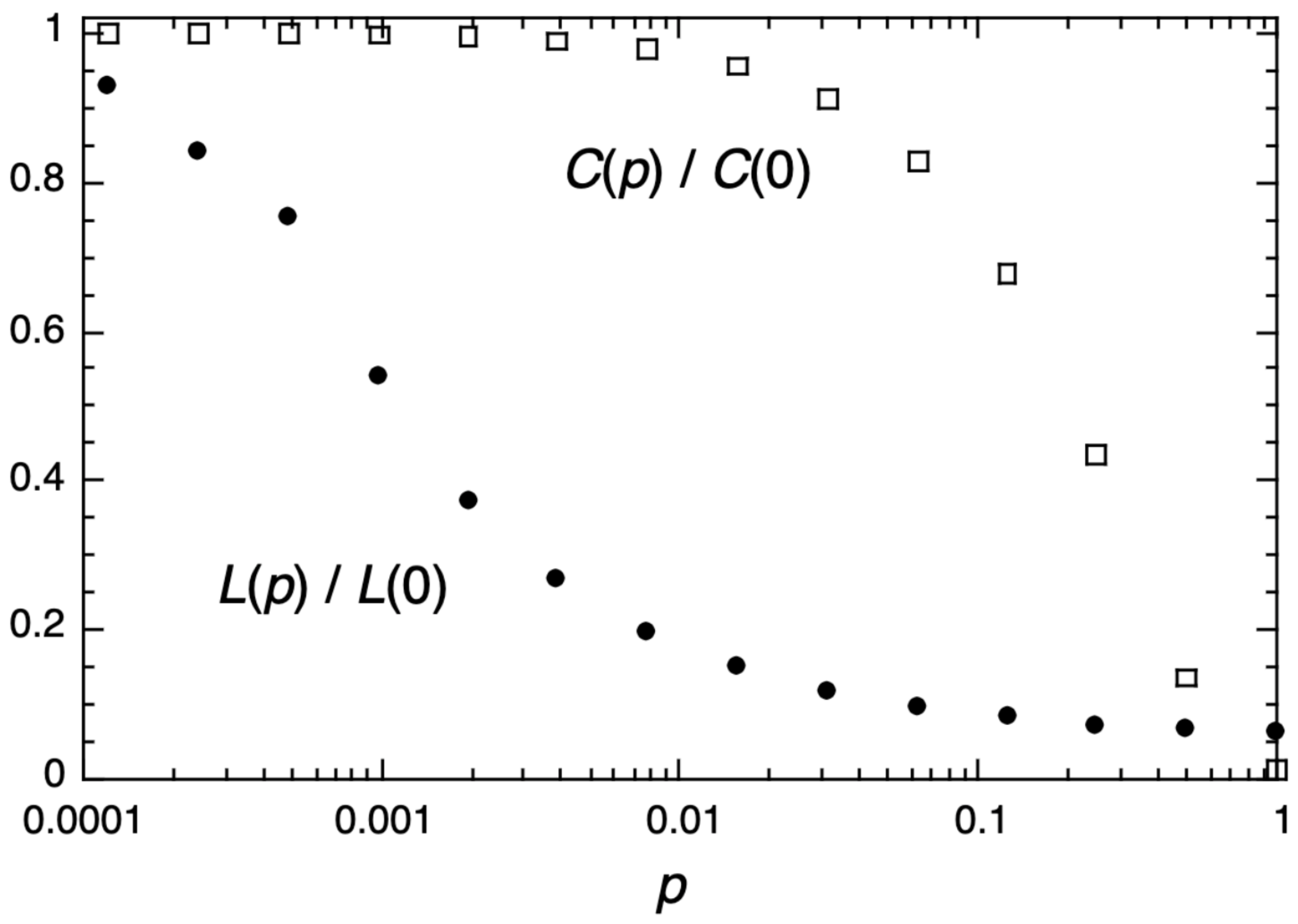
Bipartite n-node motifs



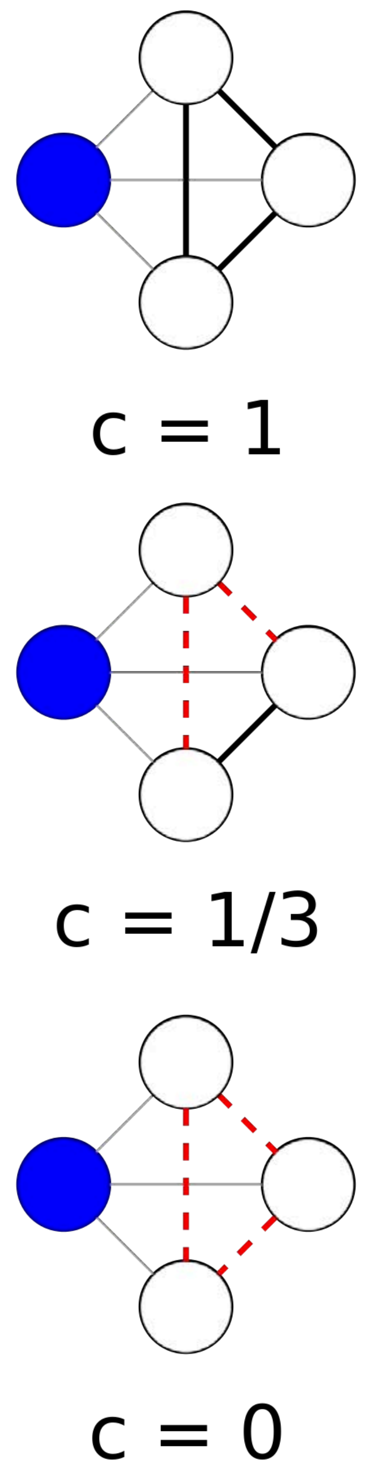
Simmons et al 2019, Oikos

Clustering coefficient - local density





L: Length of shortest path
0: values of a regular lattice



Total number of interactions between neighbors of i

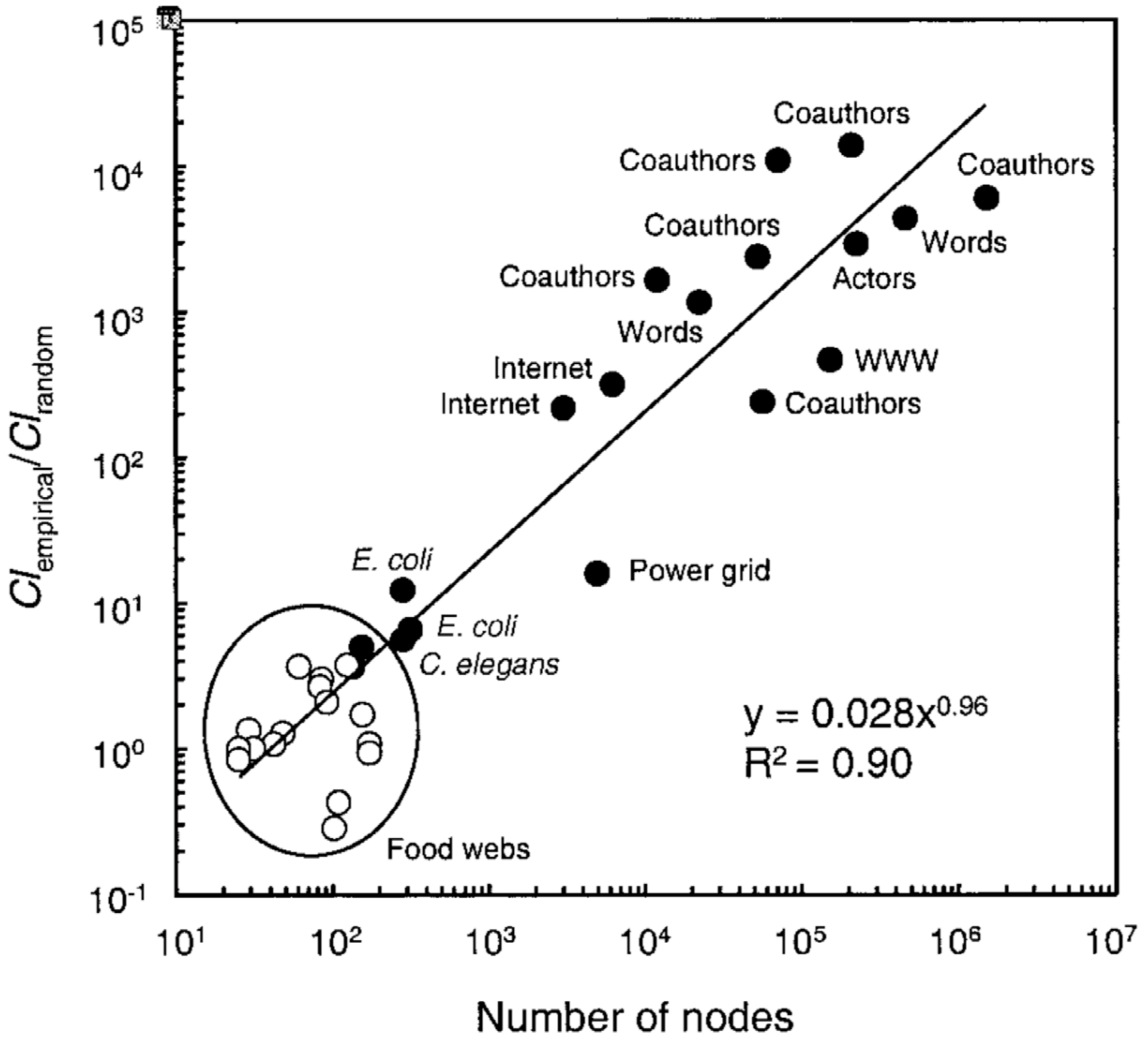
Local: $cc_i = \frac{2I_i}{k_i(k_i - 1)} = \frac{N_T}{N_{SG}}$. Low-degree nodes are more likely to have higher CC. Mean: $CC_{local} = cc_i/S$.

Global: $CC_{global} = \frac{3T_{complete}}{T}$.

Versions for directed and weighted networks exist.

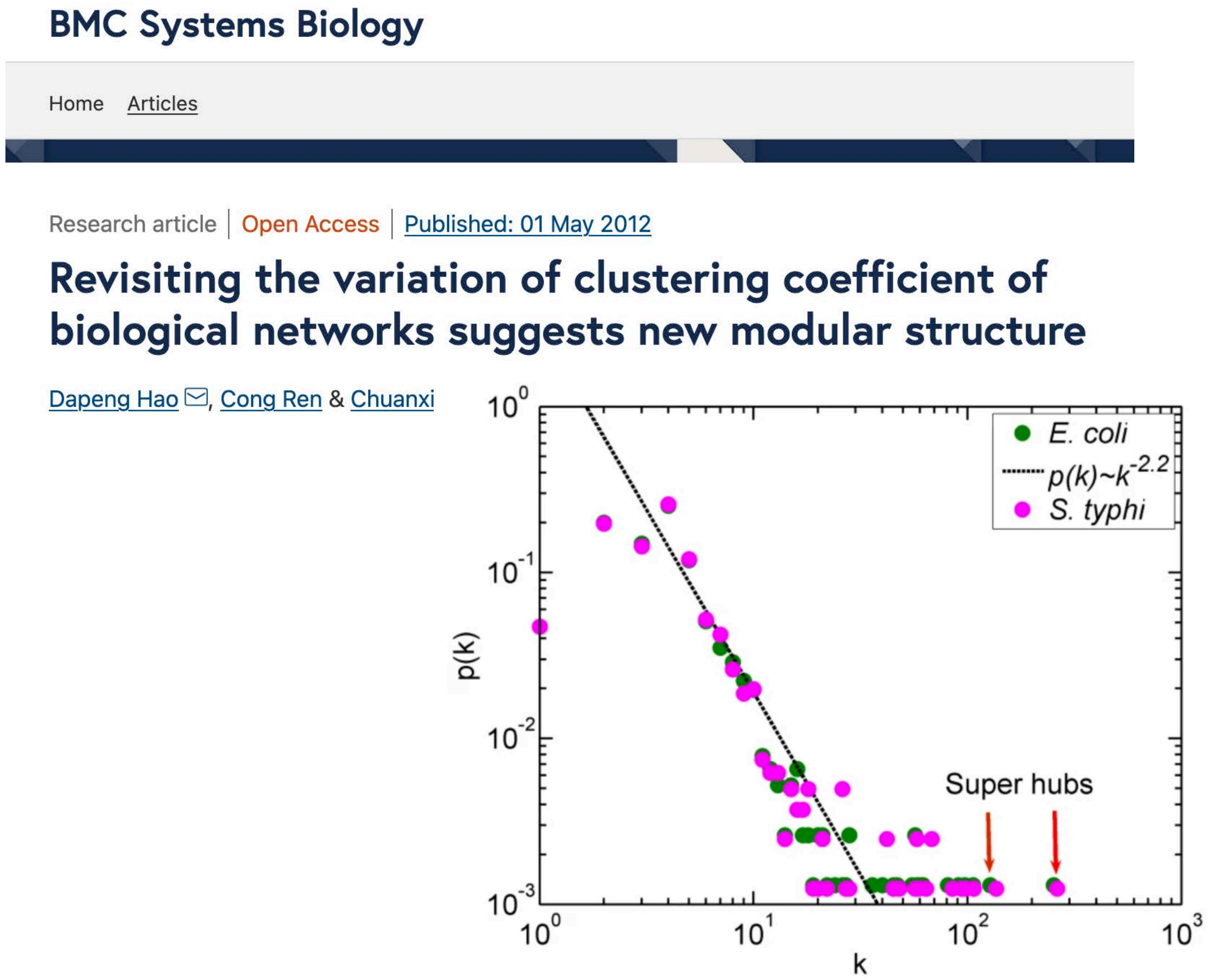
Watts & Strogartz 1998, Nature

Clustering coefficient - local density



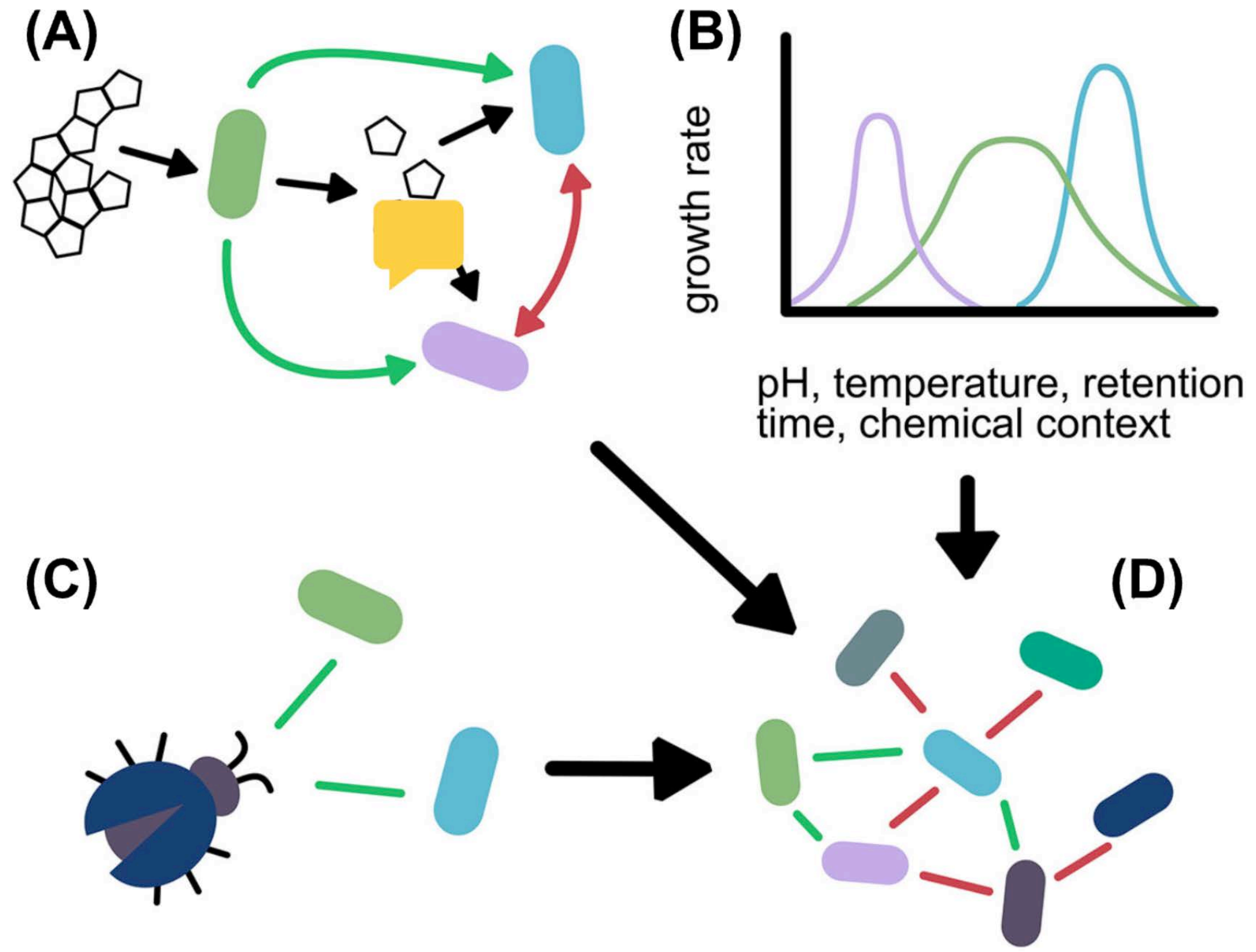
Food webs have CC close to random, but other networks have CC much higher than random.

Dunne et al 2002, PNAS

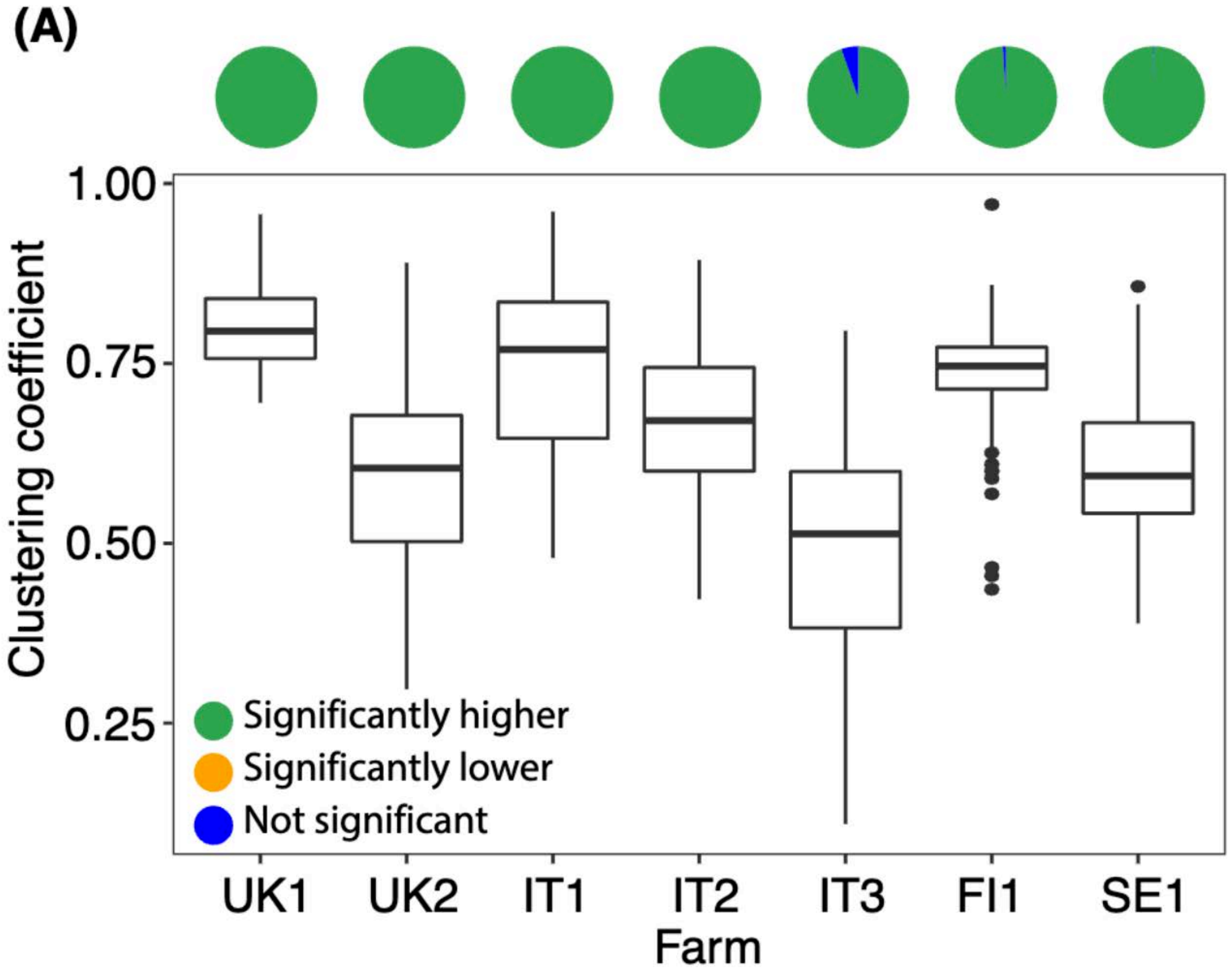


Metabolic networks: CC is affected by super-hubs

Transitivity in microbe co-occurrence



- Environmental filtering
- Cross-feeding
- Stabilized competition



What does the value of the index mean?

How to determine significance?

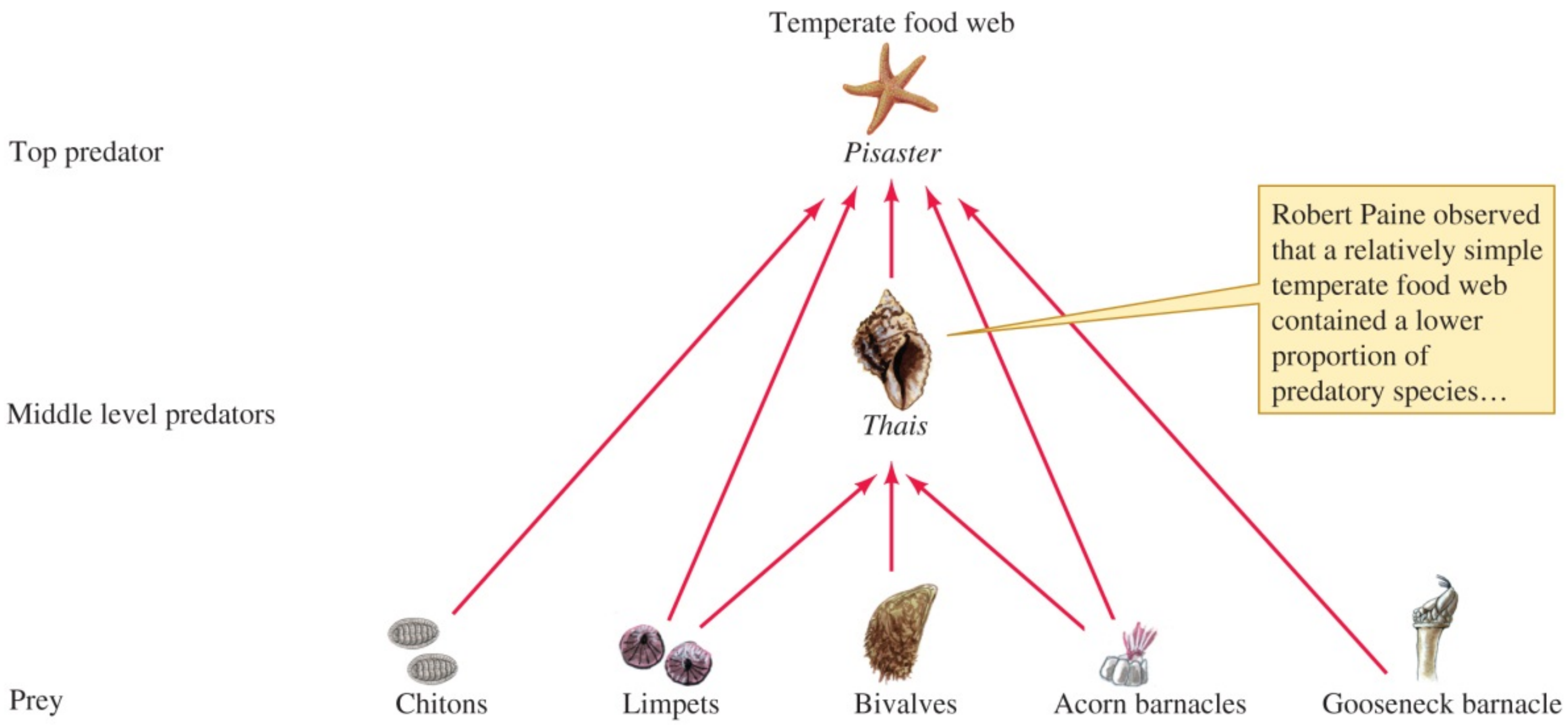
Röttjers & Faust 2018

Galai et al. 2024

How important are nodes in the network?

What is an “important” node in a:

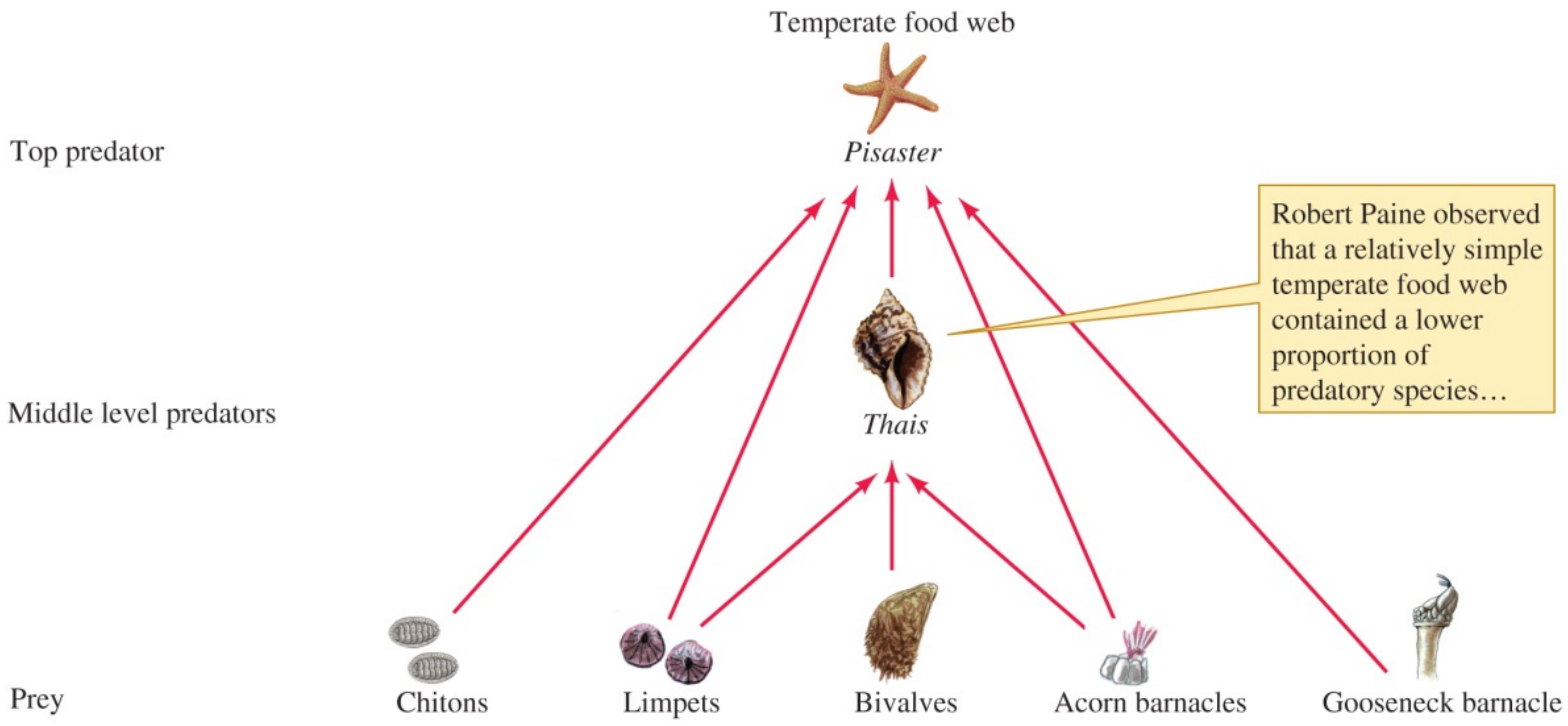
- •Food web
- •Transmission network
- •Gene-regulation network
- •Metabolic network



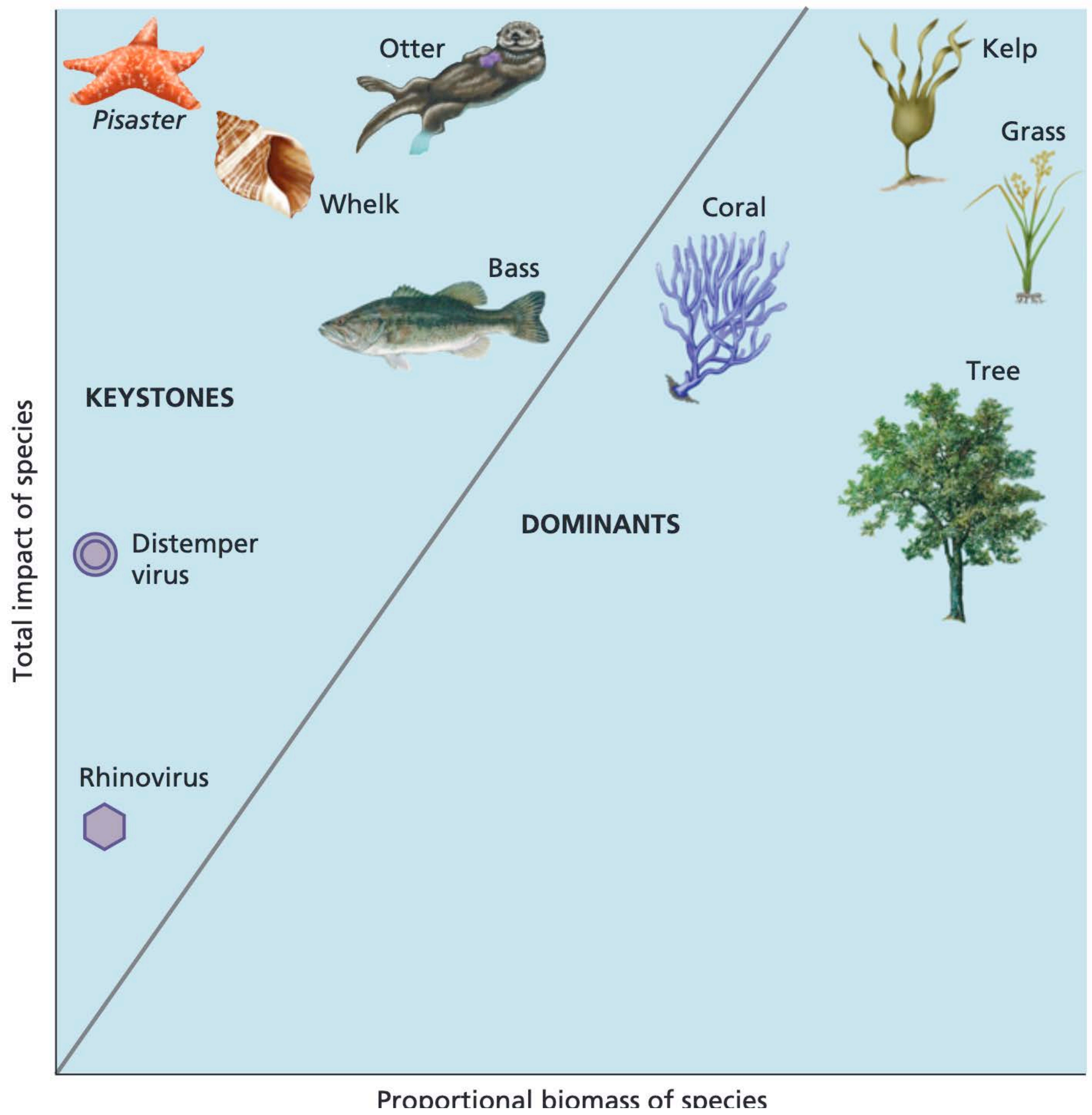
Keystones species

Predation by Pisaster star fish increases species diversity by preventing the monopolization of space by Acorn barnacles.

Robert T. Paine



Paine 1966, American Naturalist.
“Food Web Complexity and Species Diversity”



Keystone species: Disproportionate effect on the community relative to biomass.

Dominant species: Influence the community due to their high biomass.

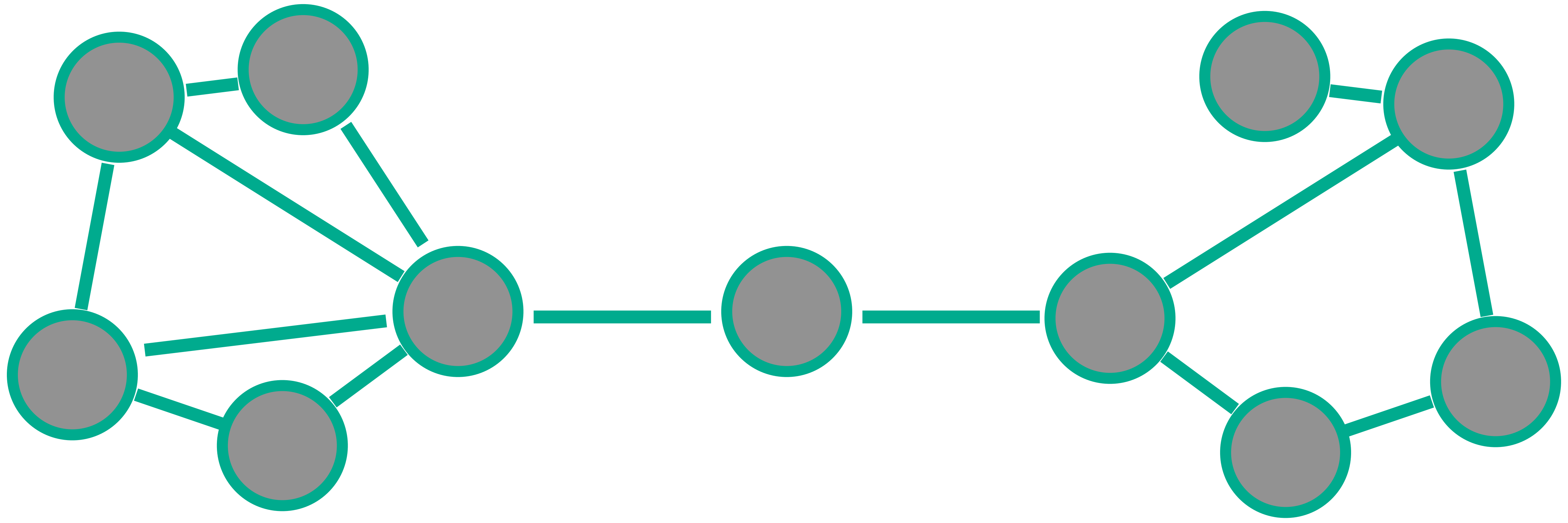
Identifying keystone species via experiments is fairly easy but predicting them a-priori is difficult. What traits does a keystone species have?

Mittelbach 2019, Community Ecology
Paine 1969, Ecology.

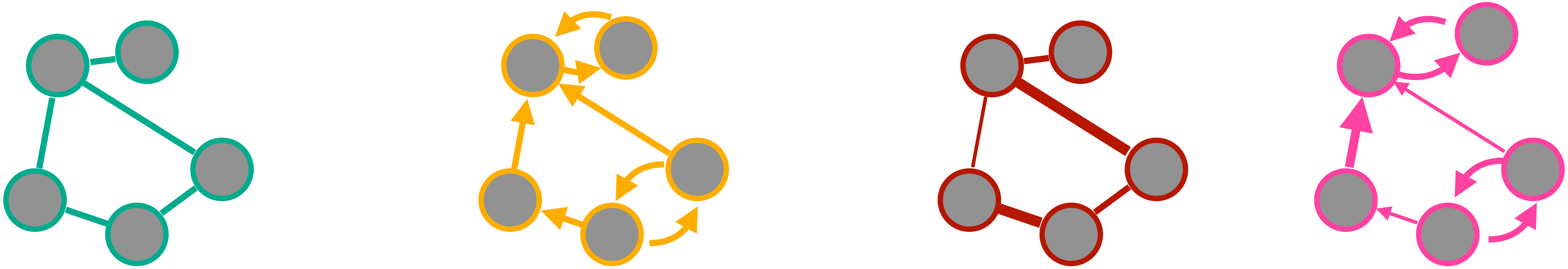
Centrality measures of ‘importance’

Standard measures:

- Degree
- Closeness
- Betweenness
- Eigenvector
- Google’s page rank



Centrality measures can identify potential keystone species but their importance needs to be validated experimentally.



Binary (unweighted) undirected

Binary (unweighted) directed

Weighted undirected

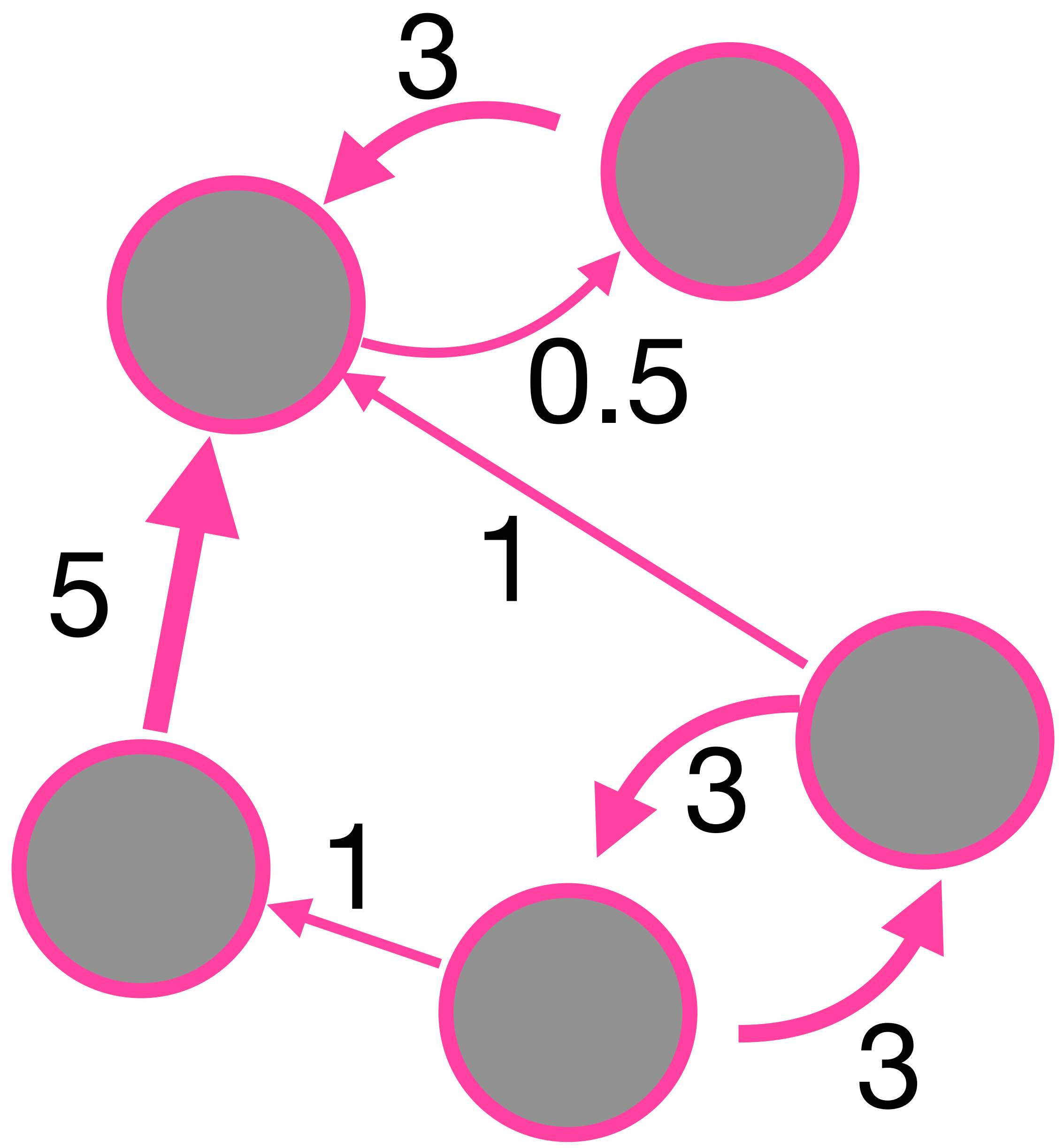
Weighted directed

Degree: number of links a node has

In-degree: number of links pointing **to** a node

Out-degree: number of links pointing **from** a node

Strength: sum of link values of a node

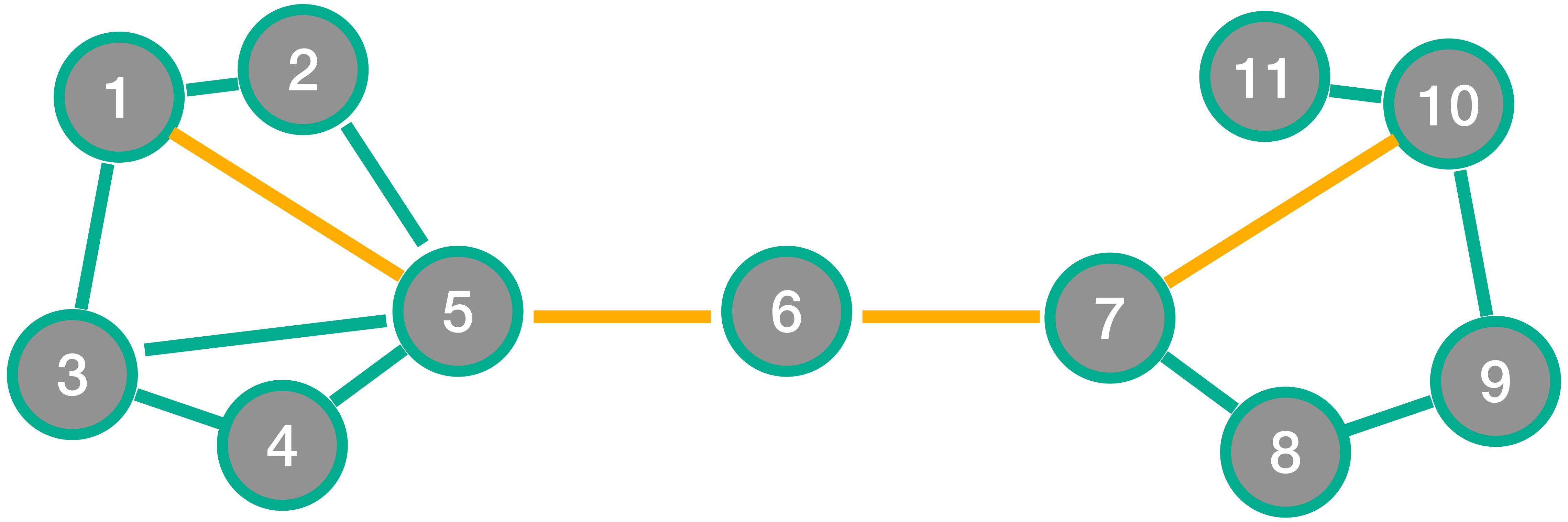


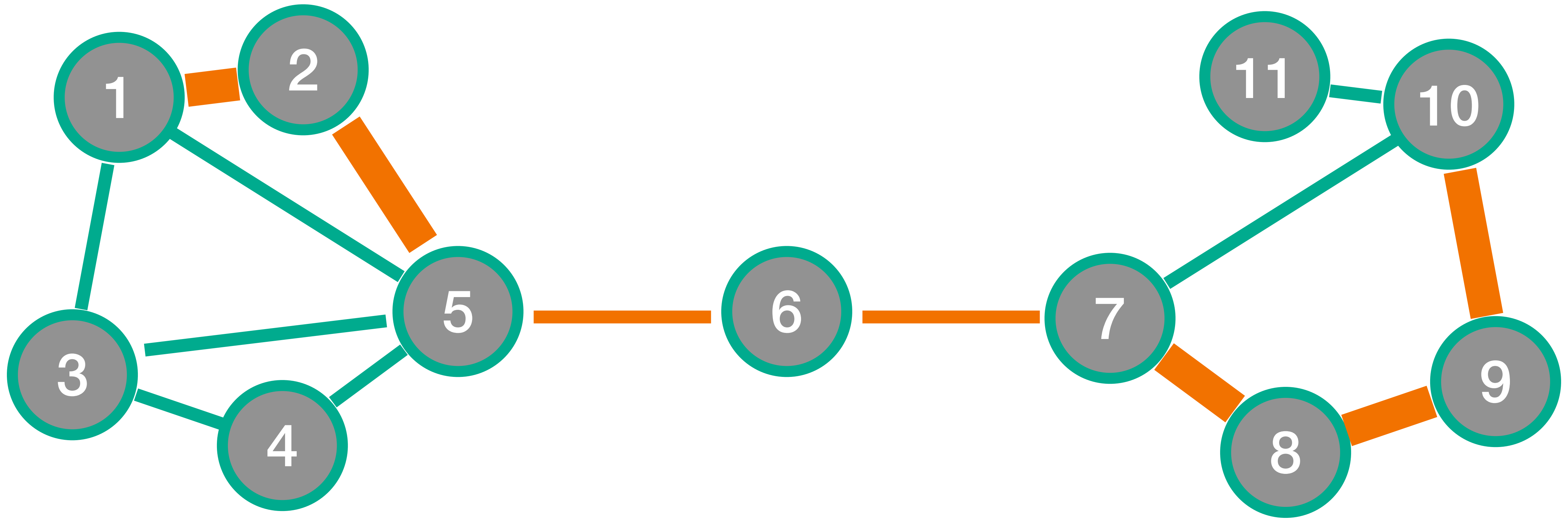
Shortest paths

A path between s and t that minimizes the sum of edge weights.

But: be careful with the meaning of edge weights!

Definitions are straightforward, algorithms (e.g., Dijkstra, BFS) not so much…





Mesoscale-level centrality measures

$$C_C(i) = \sum_{j \neq i} \frac{n-1}{d_{ij}}$$

Measures the proximity of a node to all other nodes. Node with highest Cc is closer to all other nodes

$$C_B(i) = 2 \times \sum_{j<k; i \neq j} \frac{g_{jk}(i)/g_{j,k}}{(n-1)(n-2)},$$

normalized by the number of pairs of species excluding the species under focus.

Measures the proportion of shortest paths between pairs of nodes that go through i.

$$C_E(i) = \frac{1}{\lambda} \sum_j A_{ij} C_E(j)$$

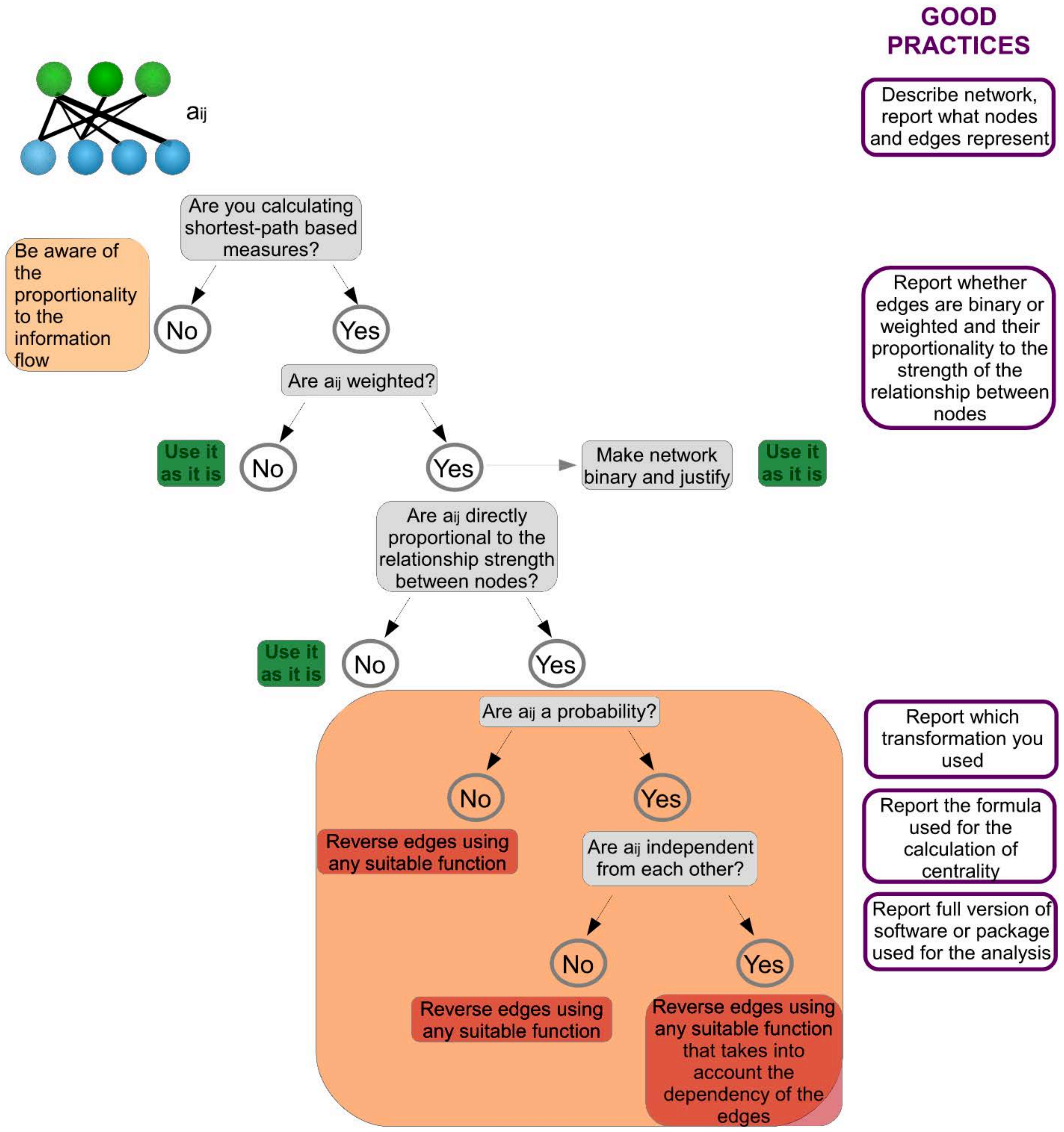
Measures flow using eigenvectors. Interaction with more influential species contributes more to the centrality than low-scoring species

Ideas for applications in different networks? Interpretations?

Shortest path algorithms typically **minimize** the value of the path between two nodes calculated as the sum of the edge weights.

Need to ensure that the edge weight is **inversely proportional** to the information flow (meaning of the edge).

- Weight reflects flow (more weight=more flow): apply algorithm as is (dispersal time, resistance).
- Weight is inverse to flow (more weight=less flow): use **inverse** weights.

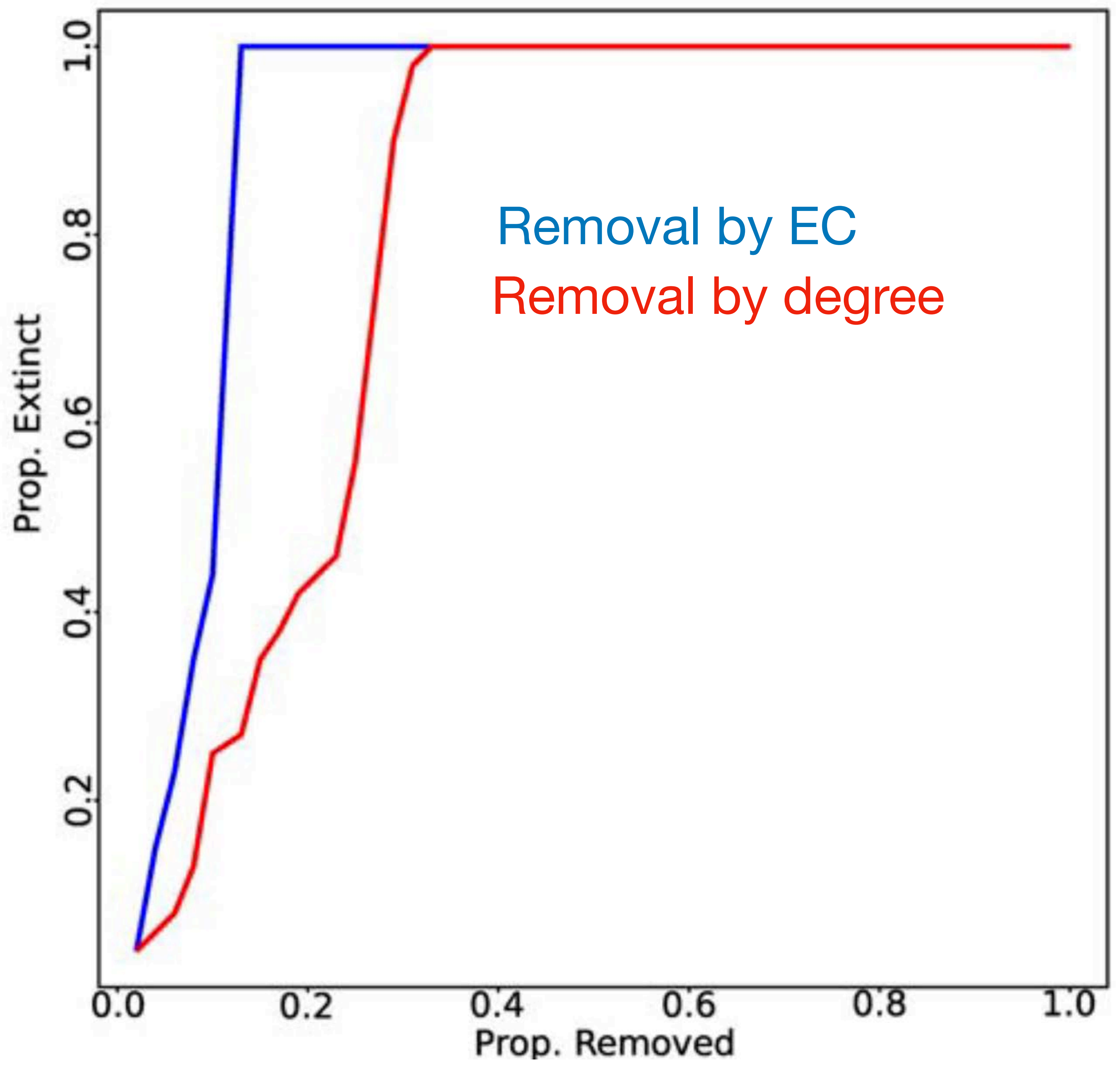


Costa et al 2019, Sci Reports

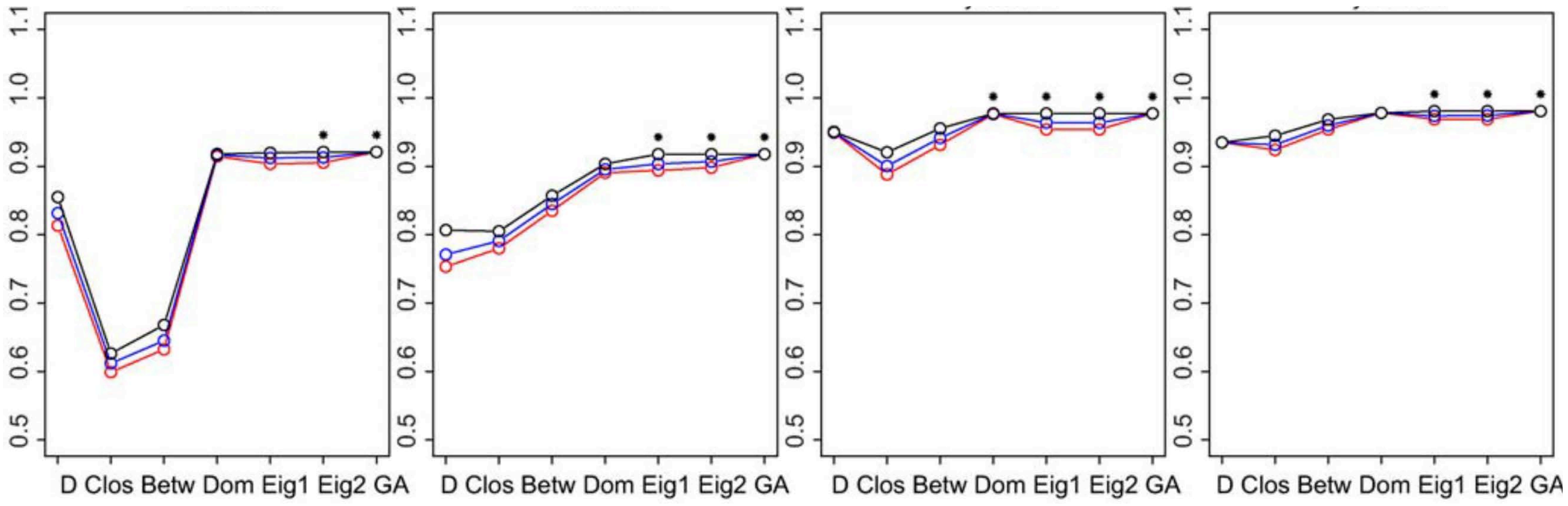
OPEN ACCESS Freely available online
PLoS COMPUTATIONAL BIOLOGY

Googling Food Webs: Can an Eigenvector Measure Species' Importance for Coextinctions?

Stefano Allesina[1]*, Mercedes Pascual[2,3,4]



Speceis position, rather than the number of links determines extinction cascasdes



Centrality measures have been heavily used in ecology (and actually everywhere…)

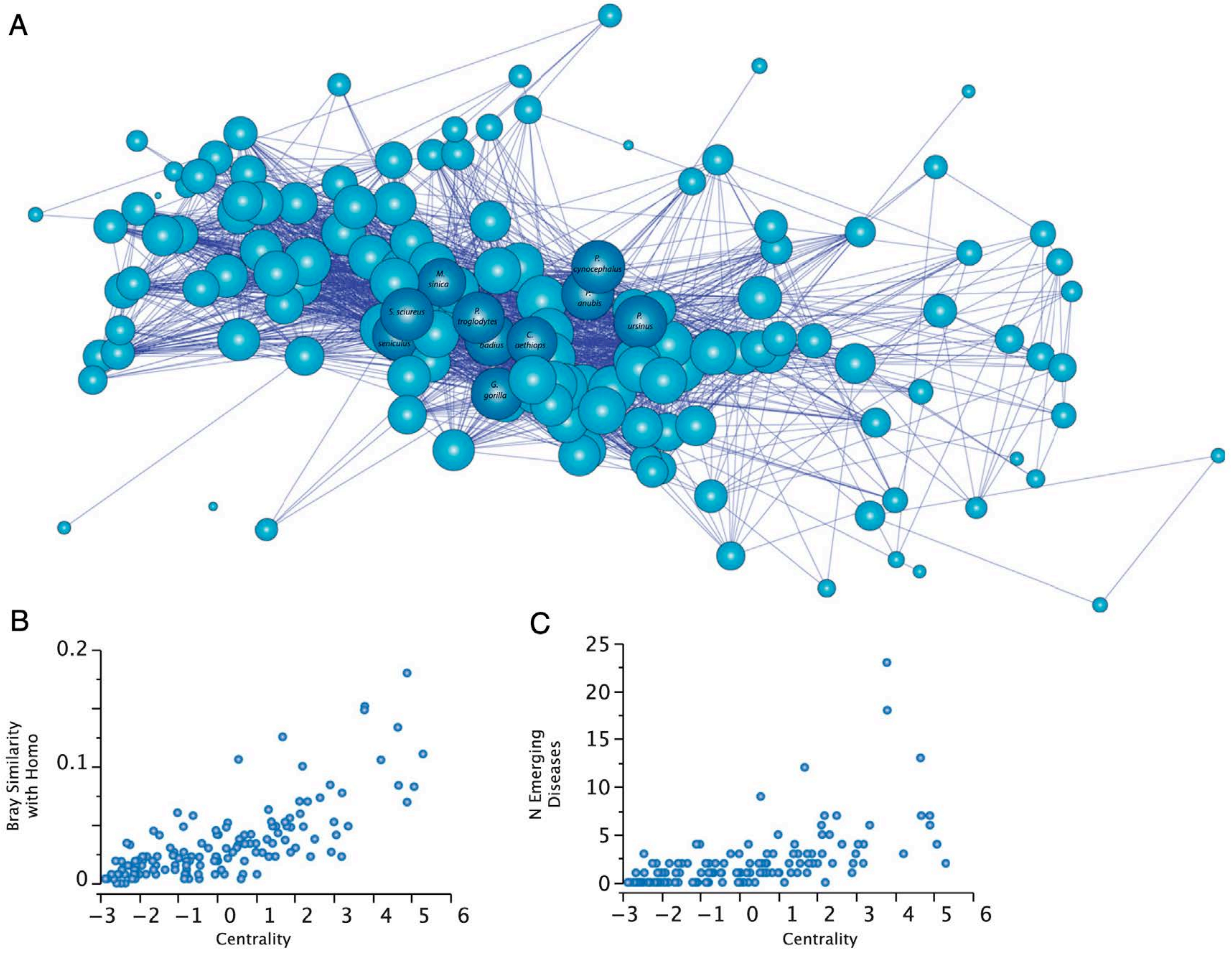


Fig. 1. Central primates shared more emerging infectious diseases with humans. (*A*) Unipartite network depicting the pattern of shared parasites by primates species. Each node represents a primate species. The links among nodes depict shared parasites (i.e., two nodes are linked whenever they share a parasite species). The network representation was generated with the Kamada-Kawai energy-minimization algorithm, which associates the mathematical centrality—calculated as the first factor of a principal component analysis of the five centrality metrics computed for each primate species—with the topological centrality in the network. Thus, nodes in the center of the presented network are more central than nodes in the periphery. The size of the nodes is proportional to the number of EIDs. Dark blue, the top 10 primates sharing more EIDs with humans. (*B*) Relationship between centrality and the similitude between human and primate parasite communities based on the Bray index. (*C*) Relationship between centrality and EiD richness found in a primate species.

Parasite sharing networks of primates.

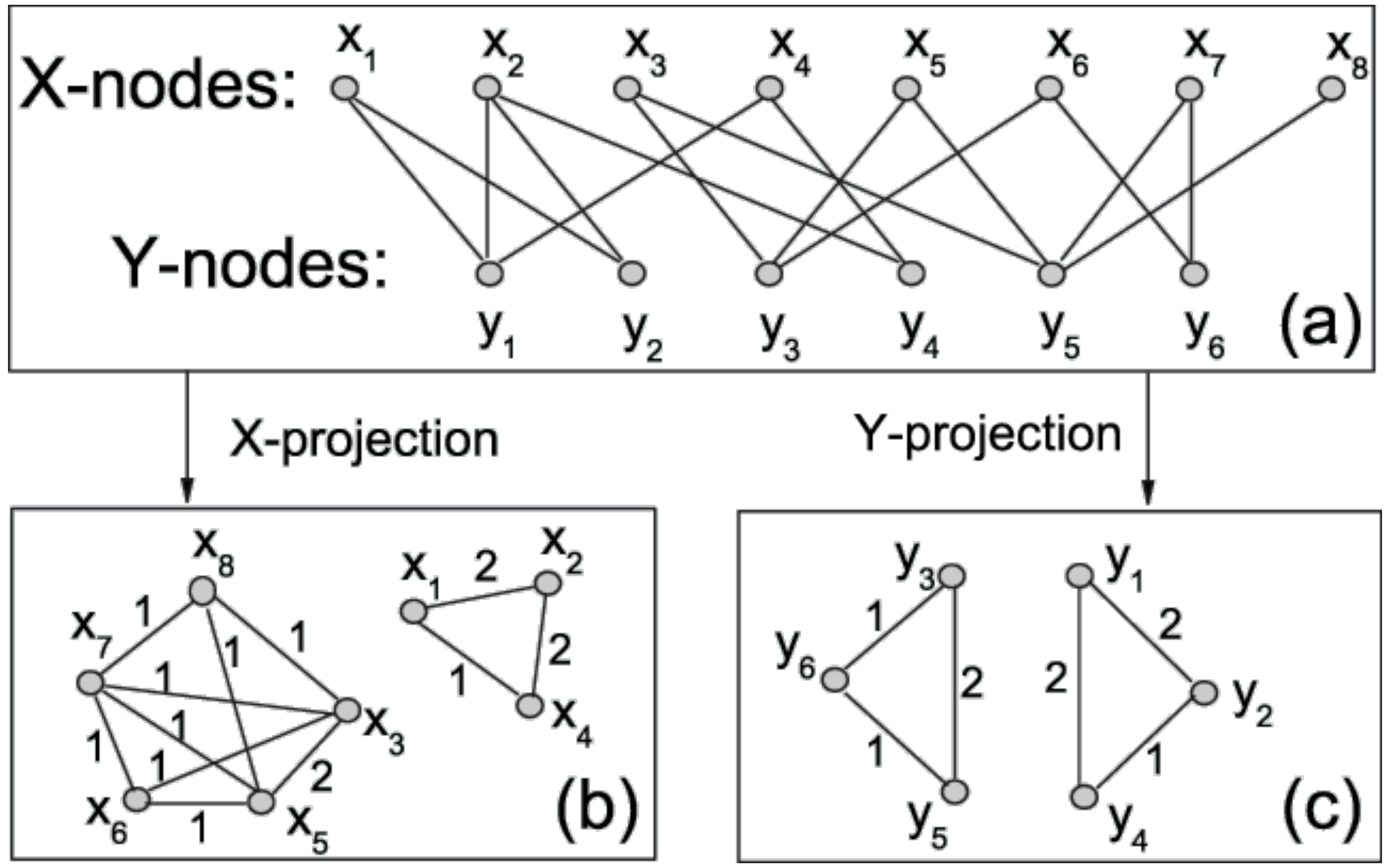
Control for sampling effort

Five centrality measures—> composite metric

Table S1. Matrix of Product-Moment correlation among the Centrality Indices. All correlations have p-values < 0.0001.

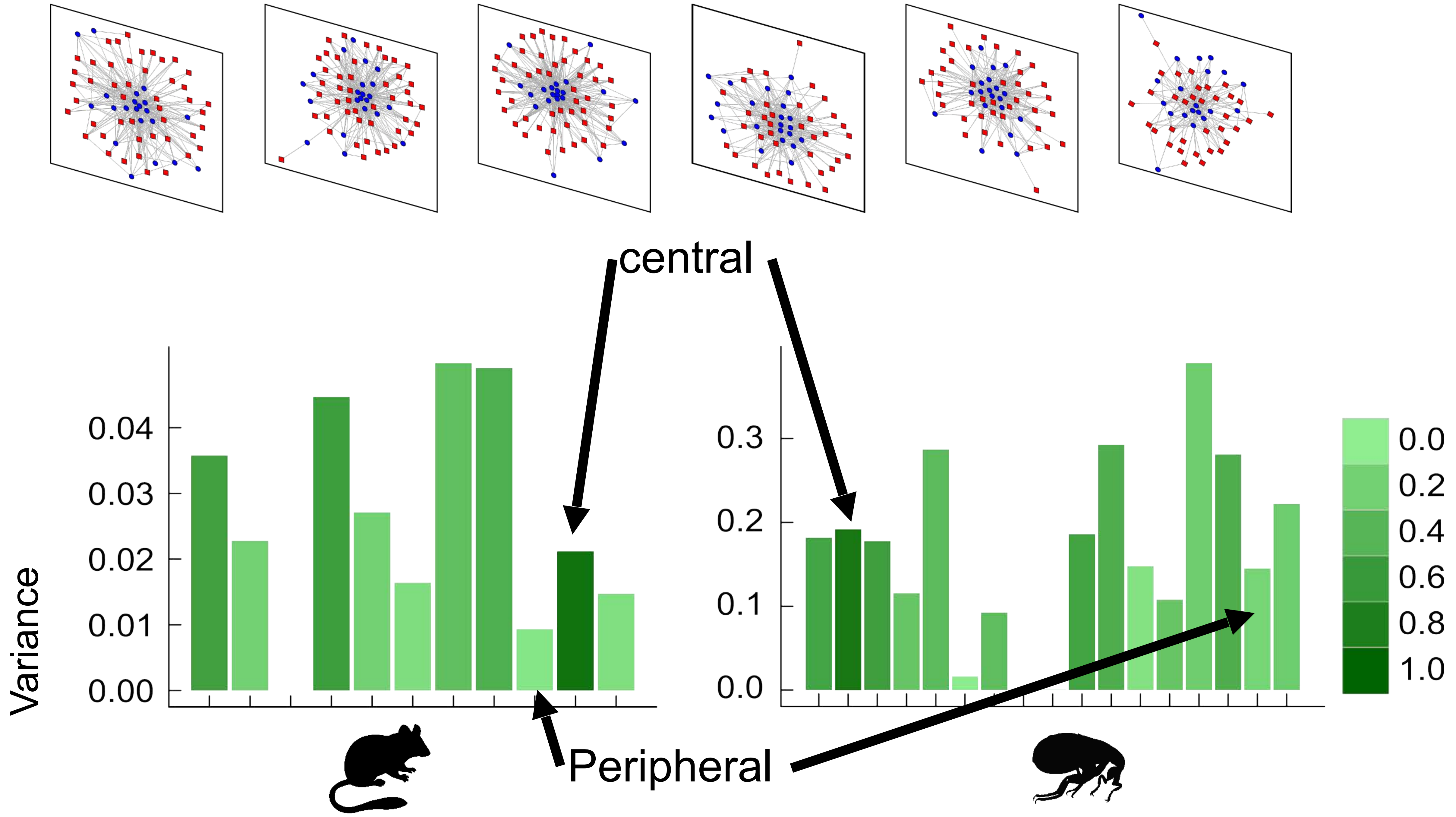
	Strength	Opsahl Degree	Eigenvector	Betweenness
Opsahl Degree	0.9981			
Eigenvector	0.9880	0.9838		
Betweenness	0.5763	0.5651	0.5449	
Closeness	0.8789	0.8888	0.8691	0.5601

Gomez el al 2013, PNAS



Zhou et al 2007, PRE

- Centrality measures are developed for unipartite networks.
- When **projecting** networks the meaning of the links changes.
- Multiple ways to weigh projected networks.

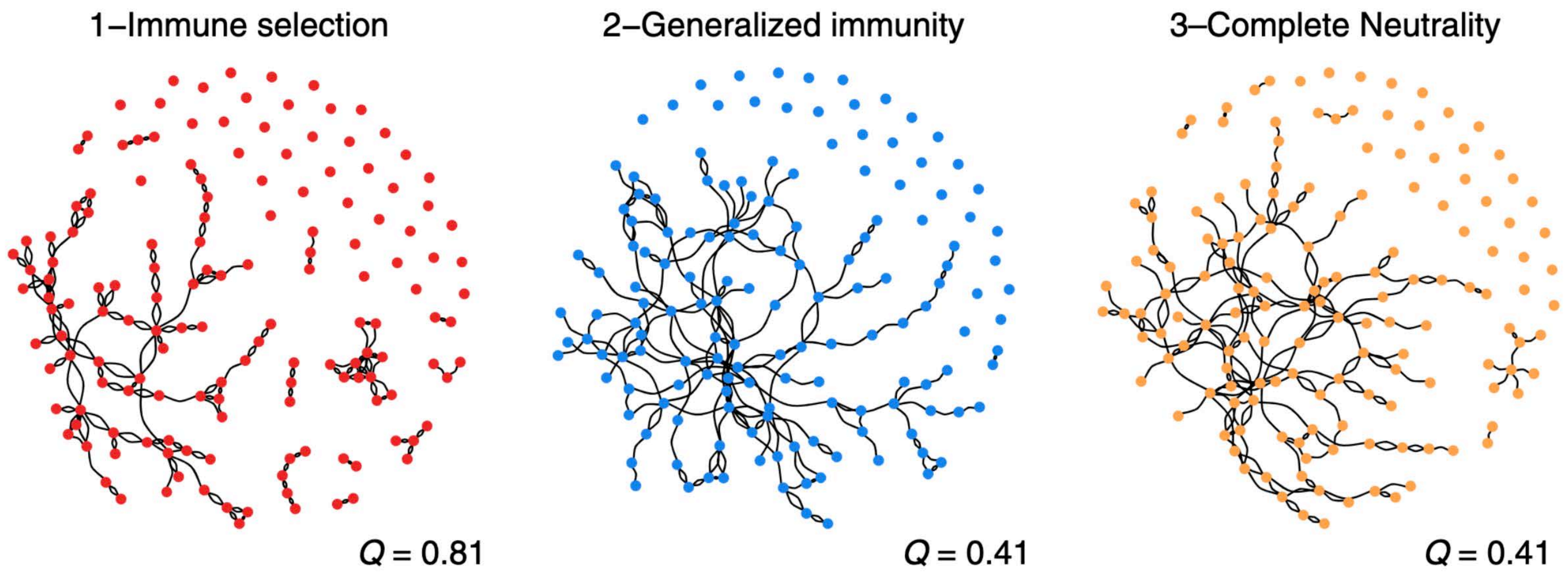


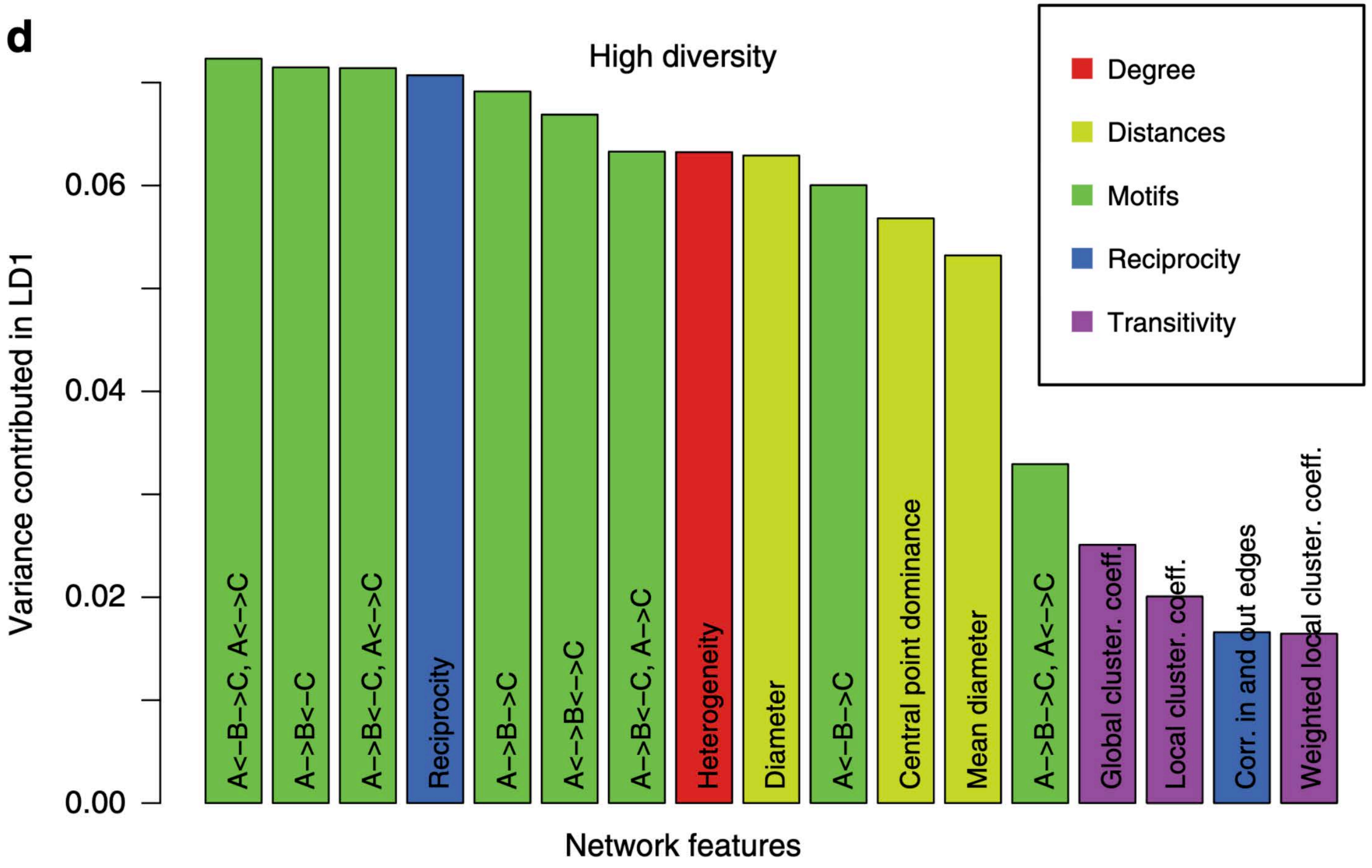
Central hosts: potentially transmit parasites to many other hosts.
Central parasites: potentially mediate the establishment of other parasites.

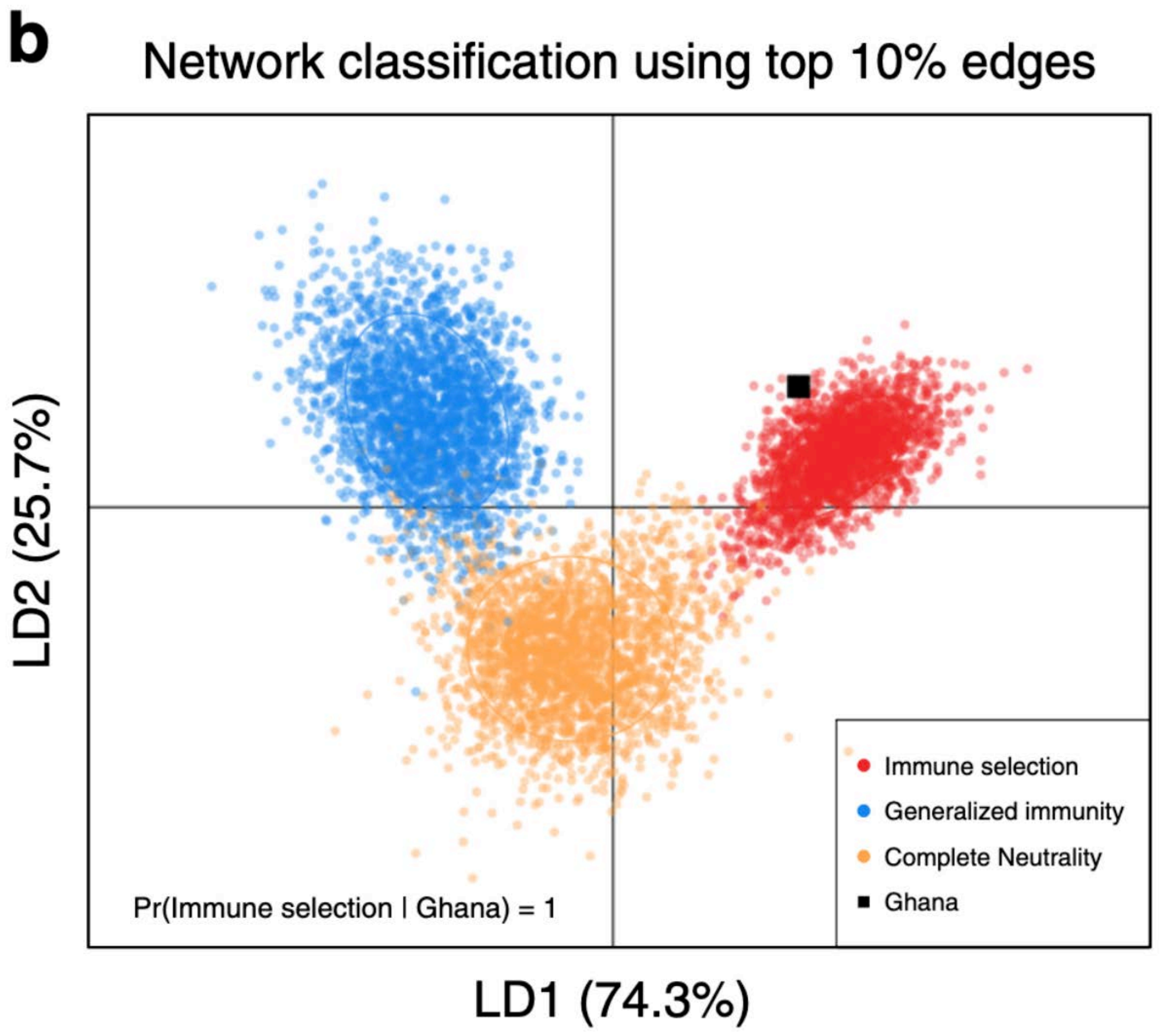
Pilosof et al 2013, J Anim Ecol

Comparing networks

- Networks with similar properties may be generated by similar processes and have similar functionality.
- Finding regularities across systems.
- Network variation in space/time/across gradients.







He, Pilosof, et al 2018, Nature Commun.

How does structure change across scales?

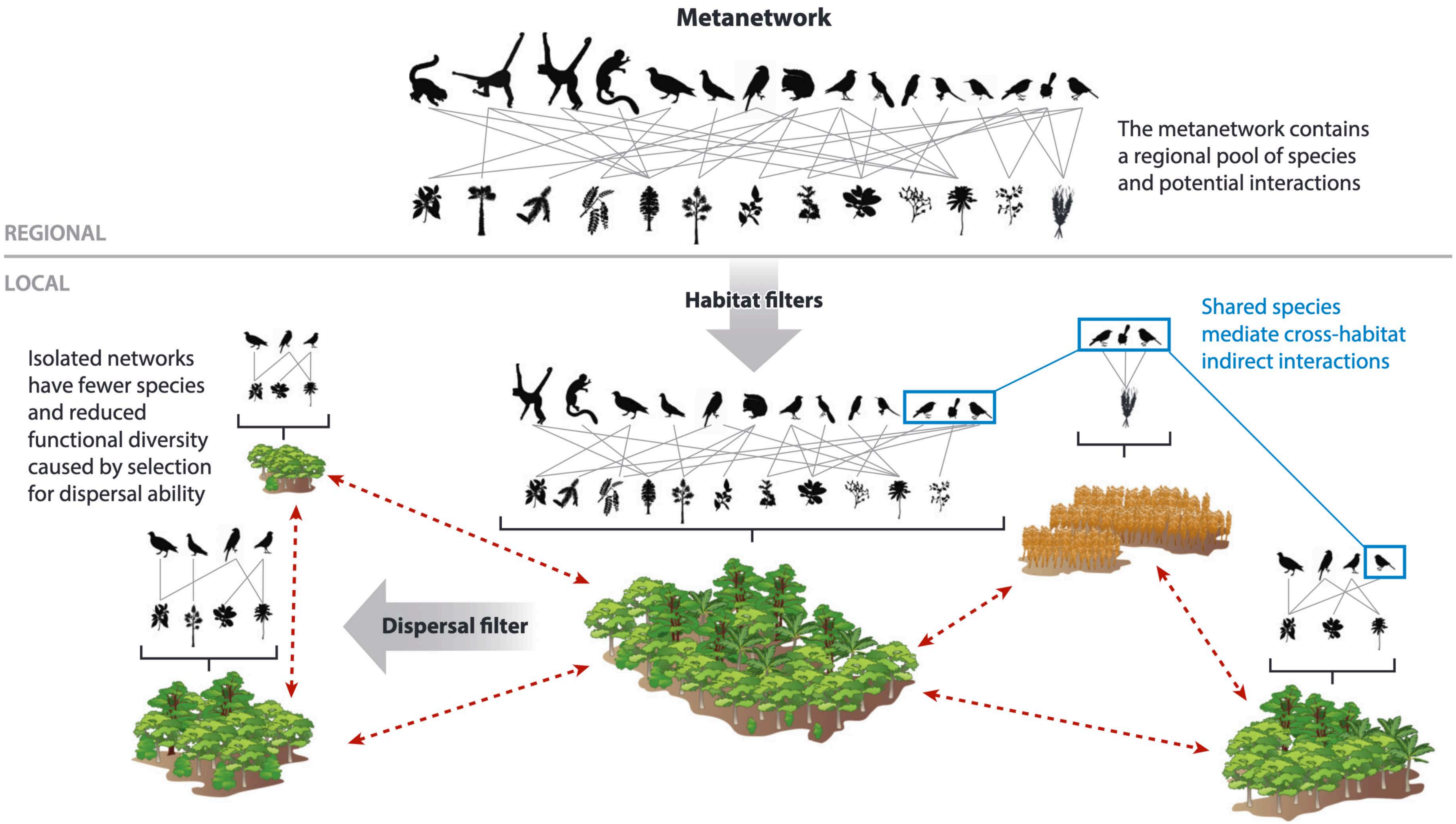
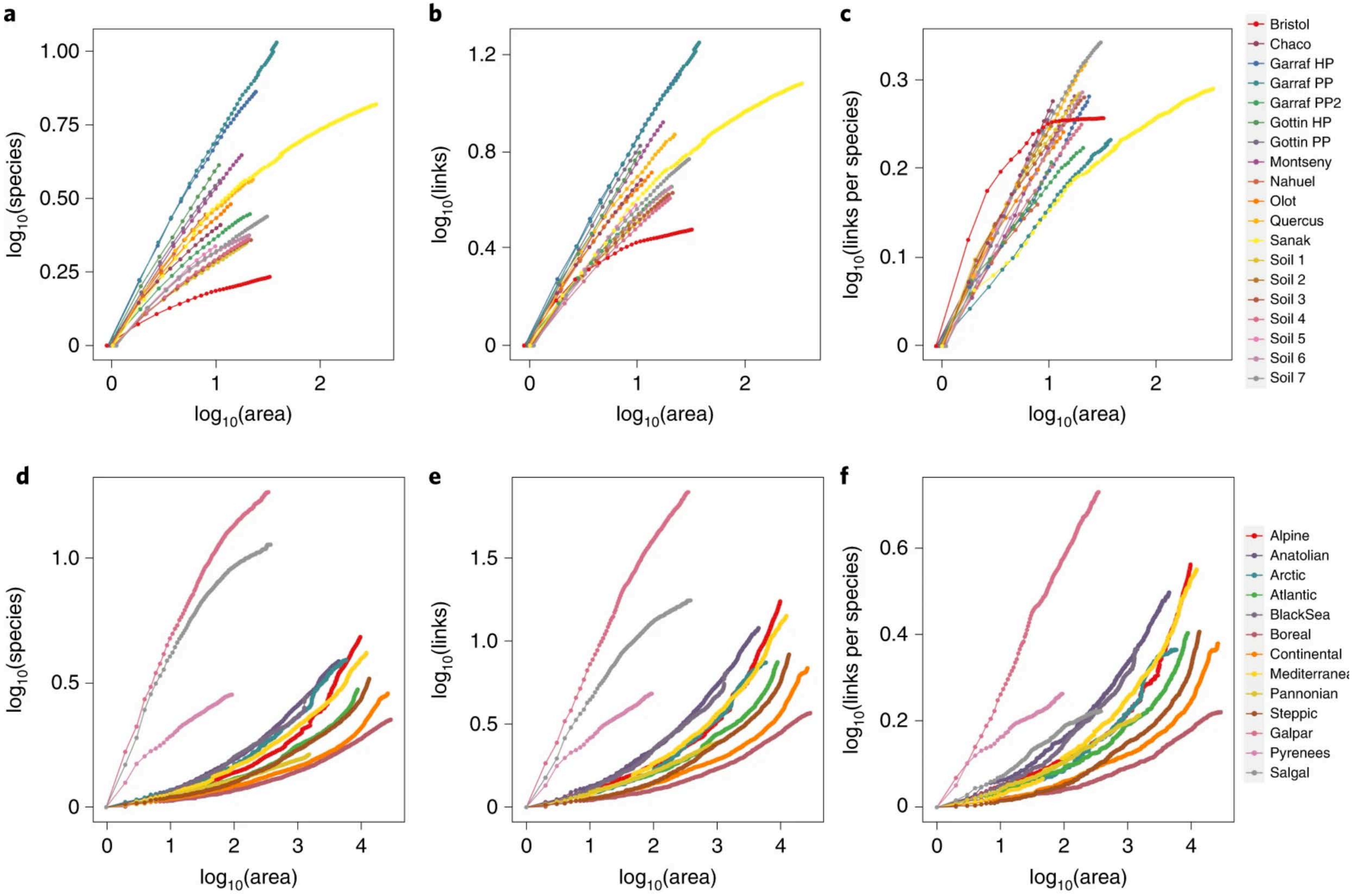


Fig. 2 from Tylianakis & Morris 2017, Annu. Rev. Ecol. Evol. Syst.

Ecological network complexity scales with area



Galiana et al 2022

Summary

We have:

- Described networks of different kinds.

- Defined multiple indices at the node and mesoscale levels:
  - Centrality, motifs, degree (and its ramifications).

- Emphasized that the selection of metrics should match the hypothesis and interpretation of results.