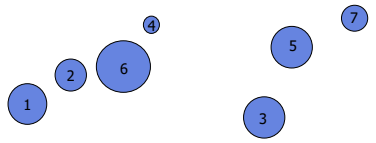
Graph Theory and Network Models in Landscape Genetics

Melanie Murphy
Associate Professor, University of Wyoming

1

What is a graph?



Object (points, **nodes**, vertices)

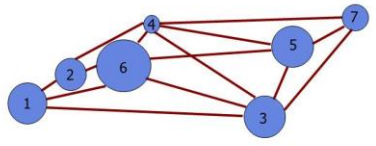
2

What is a graph?

Relationship (lines, **edges**, arcs)

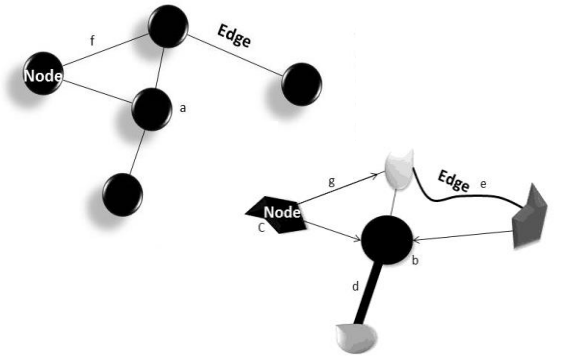
3

What is a graph?



Links indicate functional connectivity (Urban et al 2009)
Implicit and explicit relationship to metapopulations

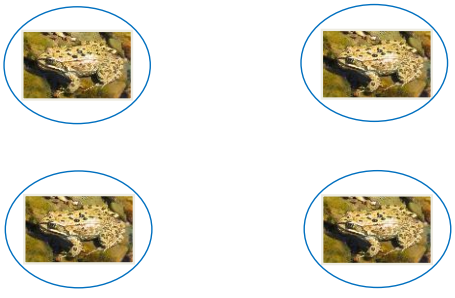
4



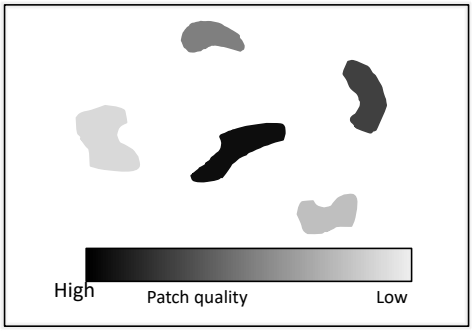
(Murphy *et al.* 2015; *Landscape Genetics: Concepts, Methods, Applications*)

5

Site Data



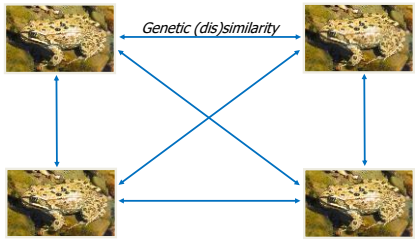
6



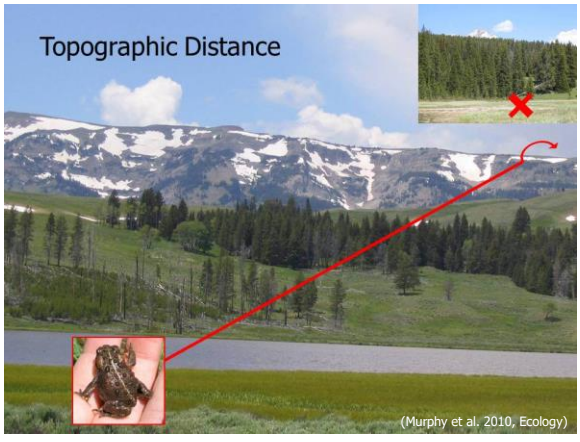
7

Pairwise data

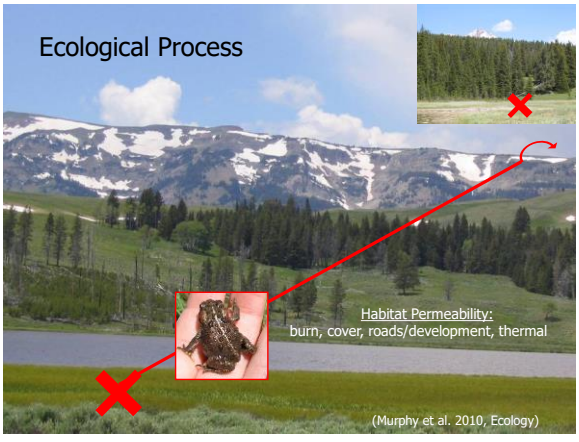
(Murphy *et al. 2015; Landscape Genetics: Concepts, Methods, Applications*, Storfer *et al.* 2007; Heredity)



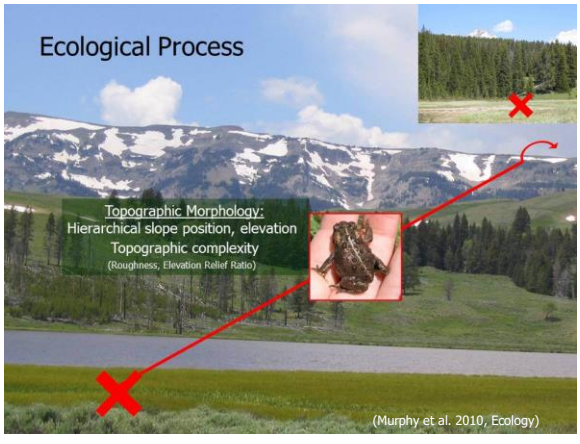
8



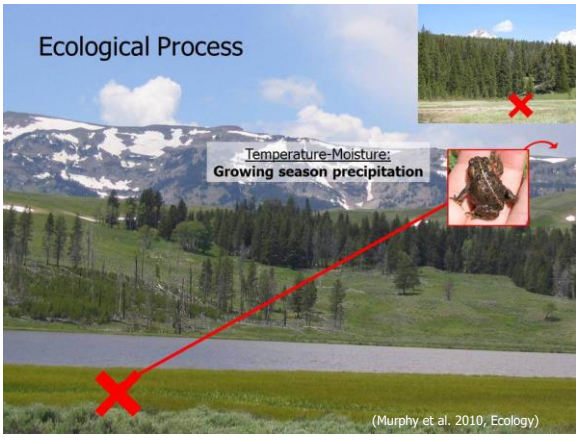
9



10

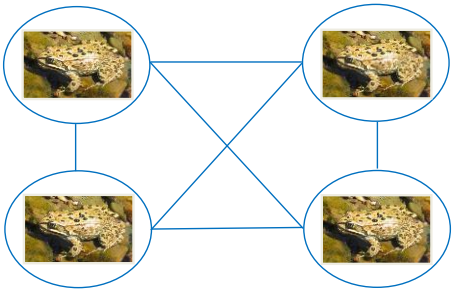


11



12

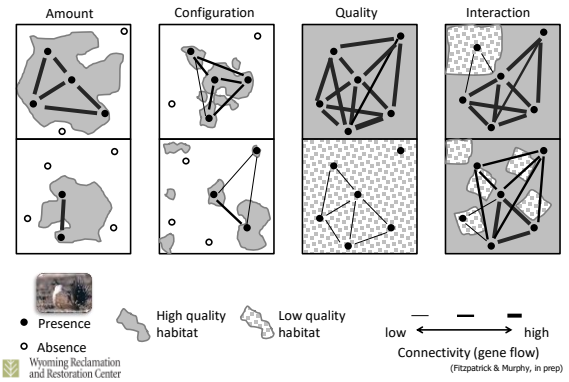
Inherently Network in Structure



(Murphy et al. 2015; *Landscape Genetics: Concepts, Methods, Applications*)

13

Functional Connectivity Hypotheses



(Fitzpatrick & Murphy, in prep)

14

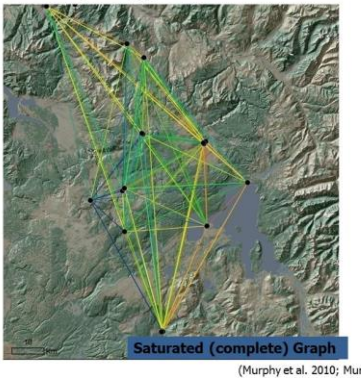
Network "Assumptions"

- Nodes easily defined/delineated
- Measuring (gene) flow
- Connections reasonable estimate of this process

(Murphy et al. 2015; *Landscape Genetics: Concepts, Methods, Applications*)

15

Network Topology – What is connected?



(Murphy et al. 2010; Murphy & Evans 2011)

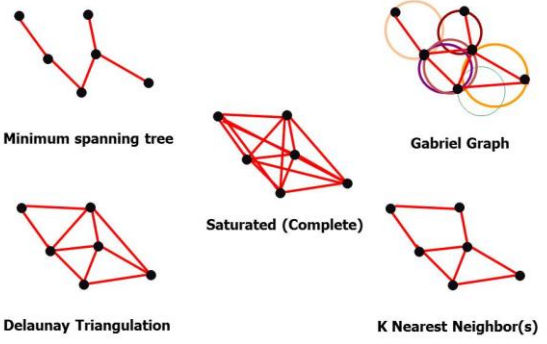
16

Network Optimization

- Reduce problem to relevant edges
- Identify connections with highest gene flow
- Can avoid overlapping connections
- Avoid long edges

17

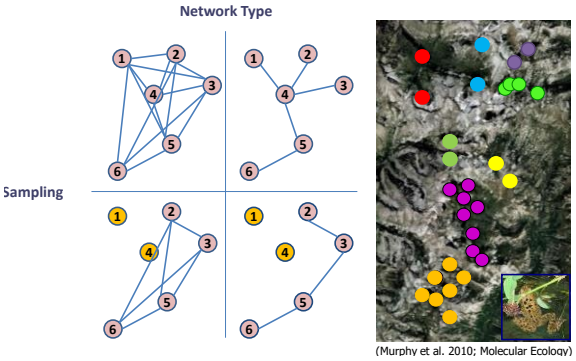
Rule-based networks



18

Sensitivity Analysis of Spatial Genetic Networks (DGS 2010)

(Naujokaitis-Lewis et al. 2013; Conservation Genetics)



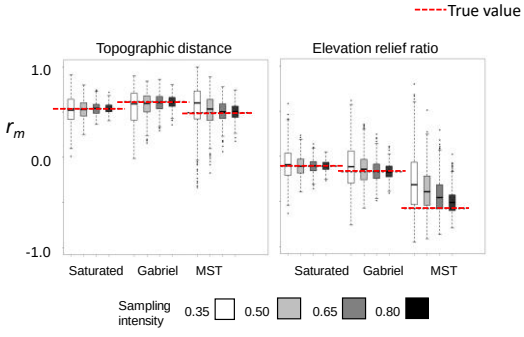
(Murphy et al. 2010; Molecular Ecology)

19

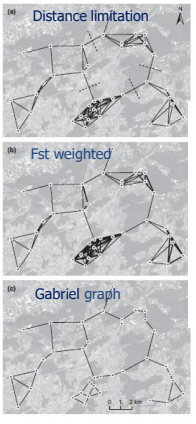
Sampling Intensity	Saturated	Gabriel	MST
Complete census ▪ True value ▪ n = 30			
50 % sampling intensity ▪ n = 15			

20

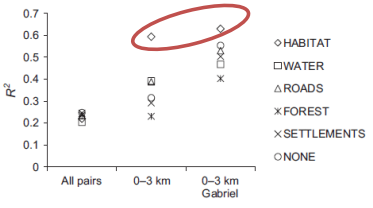
Results for Graph Function: Sensitivity of Landscape Variables



21



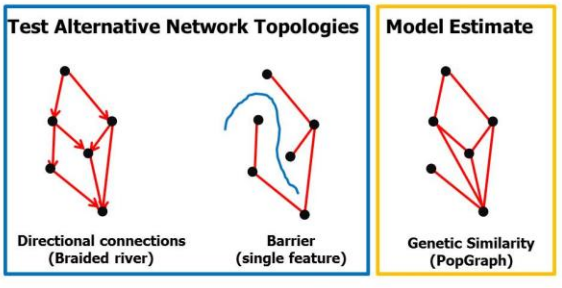
Landscape Genetics of Wetland Grasshopper



(Keller et al. 2013, Molecular Ecology)

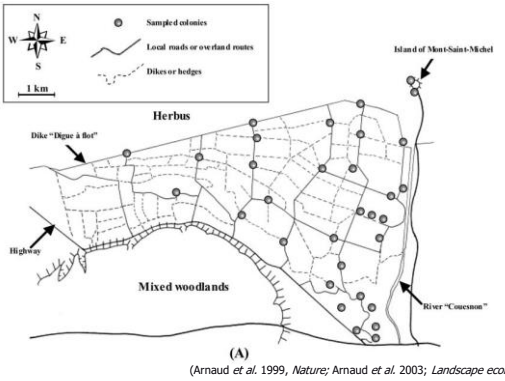
22

Hypothesized or model based



23

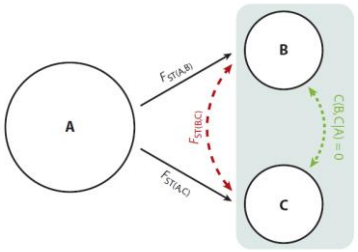
Hypothesis Based Network



(Arnaud *et al.* 1999, *Nature;* Arnaud *et al.* 2003; *Landscape ecology*)

24

Model-Based Networks: Population Graphs

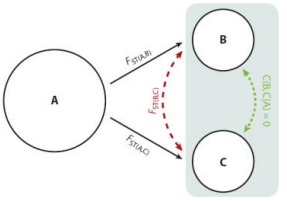


(Dyer and Nason 2004; Dyer 2015)

25

Model-Based Networks: Population Graphs

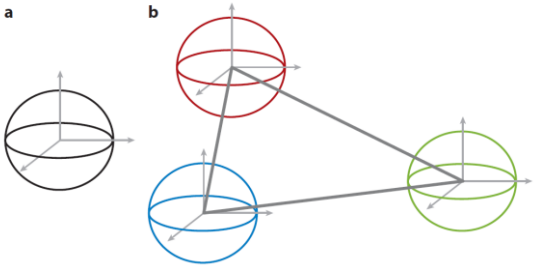
- Prune edges not adding to structure
- Nodes scaled to genetic variation
- Edges scaled to genetic distance
- Relatively assumption free



(Dyer and Nason 2004; Dyer 2015)

26

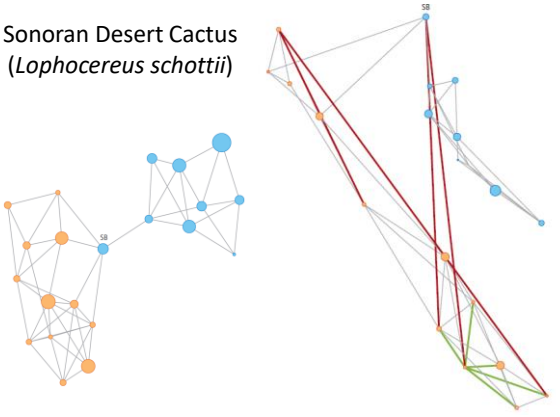
Population Graphs



(Dyer and Nason 2004; Dyer 2015)

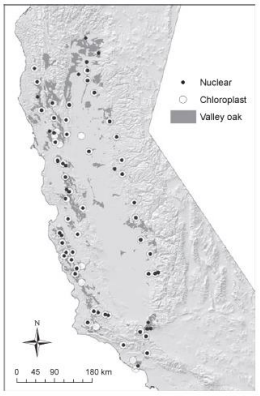
27

Sonoran Desert Cactus (*Lophocereus schottii*)



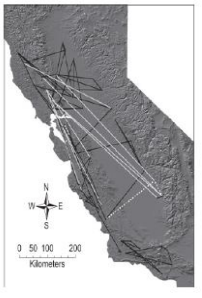
(Dyer and Nason 2004, Molecular Ecology; Dyer 2015 *Annual Review of Ecology, Evolution, and Systematics)*

28

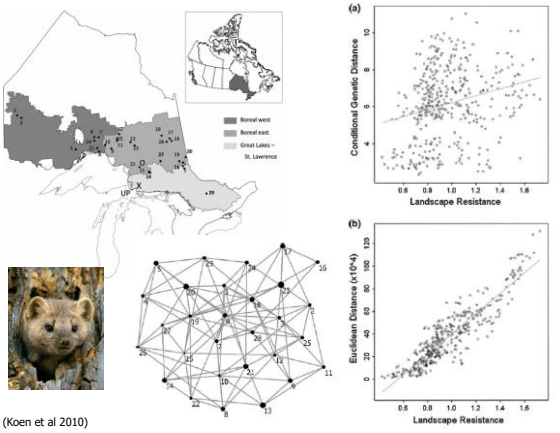


Valley Oak

(Sork et al. 2010)



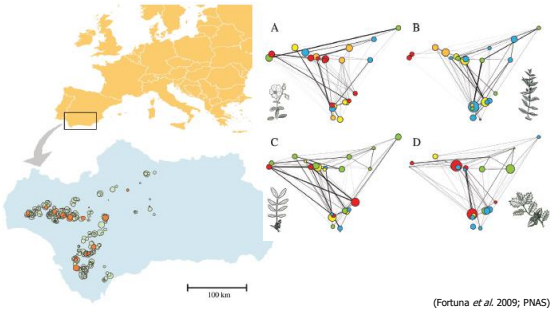
29



(Koen et al 2010)

30

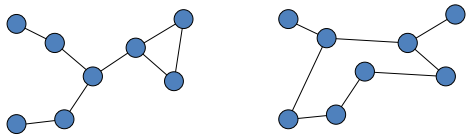
Cross-species graph comparison



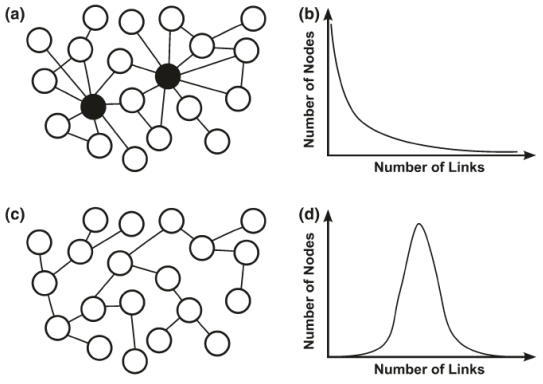
(Fortuna *et al.* 2009; PNAS)

31

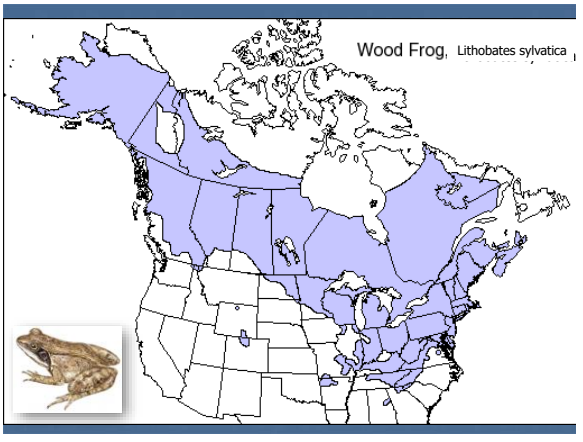
Graph Metrics



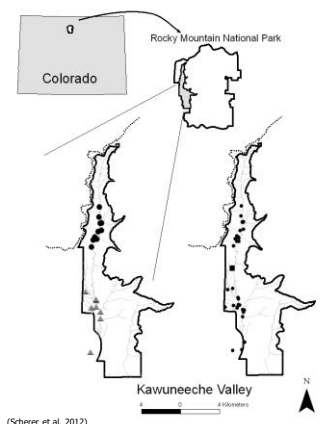
32

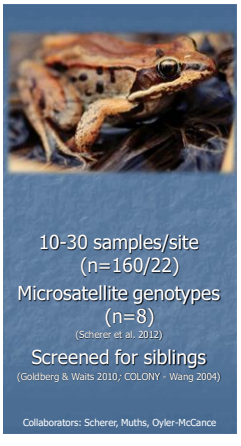


33

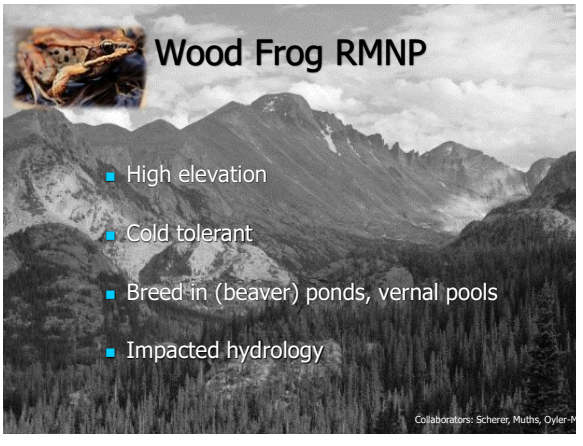


34

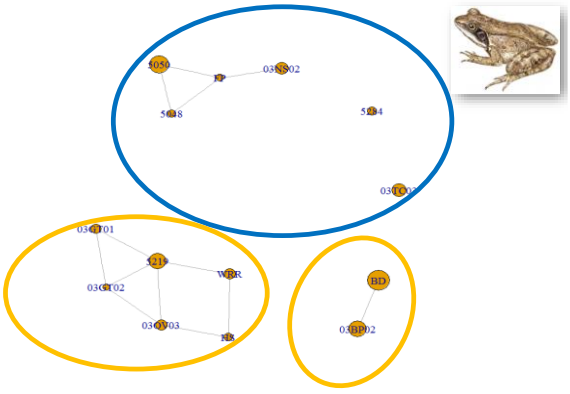




35



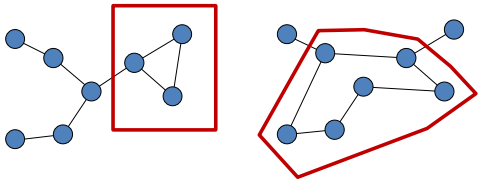
36



37

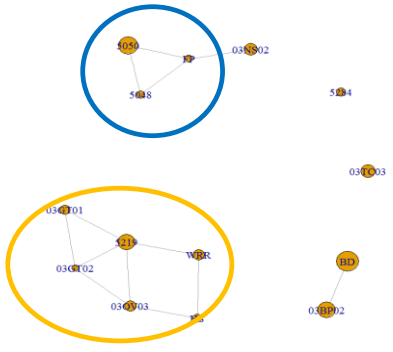
38

Graph Metrics



Cycle

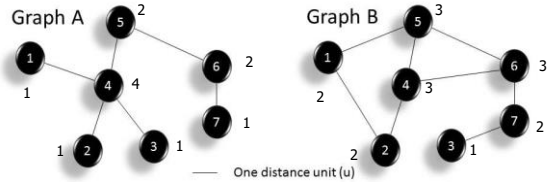
39



40

Degree

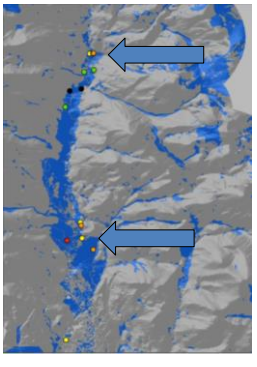
Number of nodes linked to a node



(Chapter 10, Murphy *et al.* 2015; *Landscape Genetics: Concepts, Methods, Applications*)

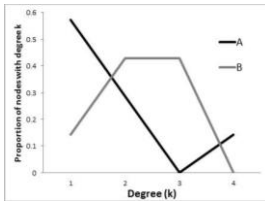
41

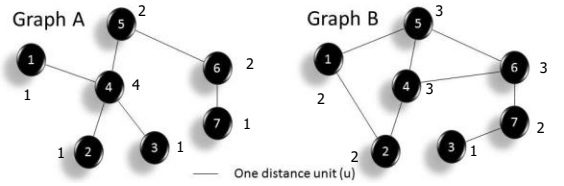
Degree



42

Degree Distribution

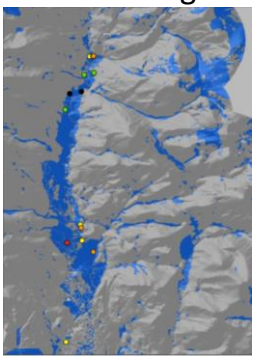


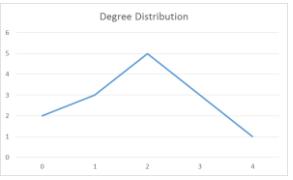


(Murphy *et al.* 2015; *Landscape Genetics: Concepts, Methods, Applications*)

43

Degree Distribution

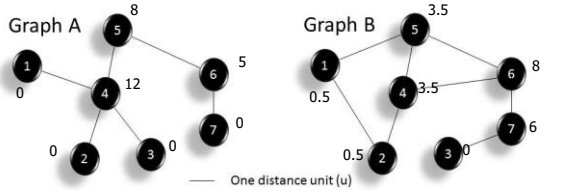




44

Betweenness

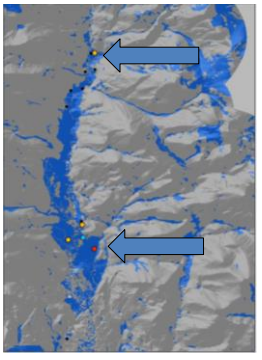
Number of times node is the shortest path



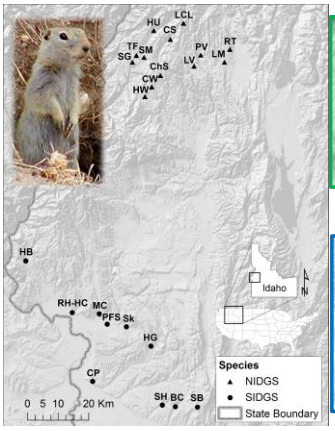
(Murphy *et al.* 2015; *Landscape Genetics: Concepts, Methods, Applications*)

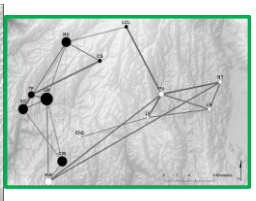
45

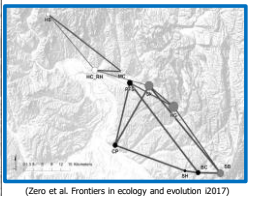
Betweenness



46



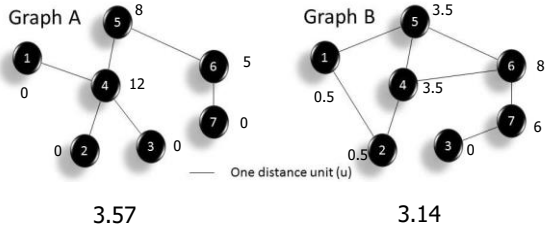




(Zero et al. Frontiers in ecology and evolution i2017)

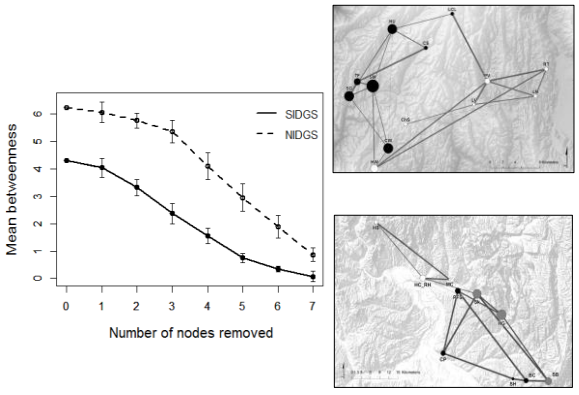
47

Mean Betweenness



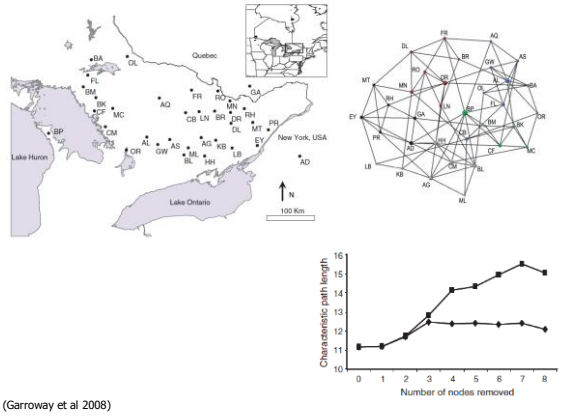
(Murphy *et al.* 2015; *Landscape Genetics: Concepts, Methods, Applications*)

48



(Zero et al. 2017)

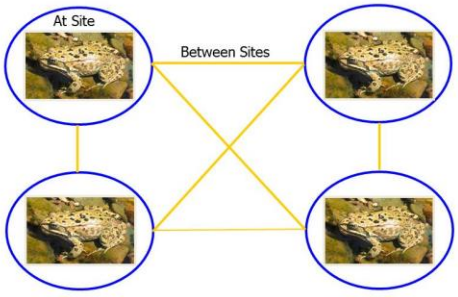
49



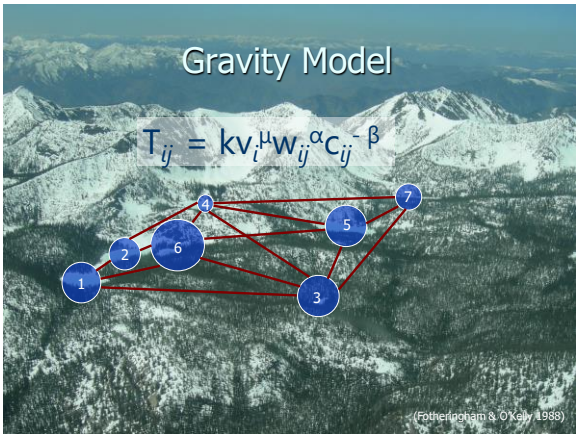
(Garroway et al 2008)

50

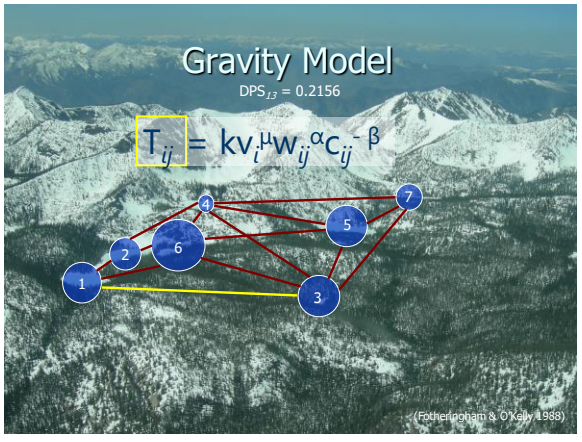
Do at site and between site processes limit connectivity?



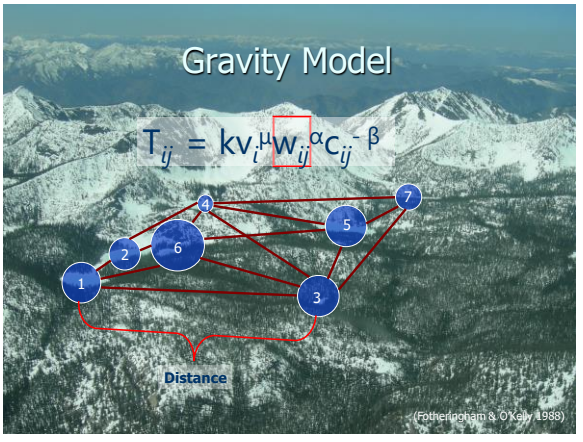
51



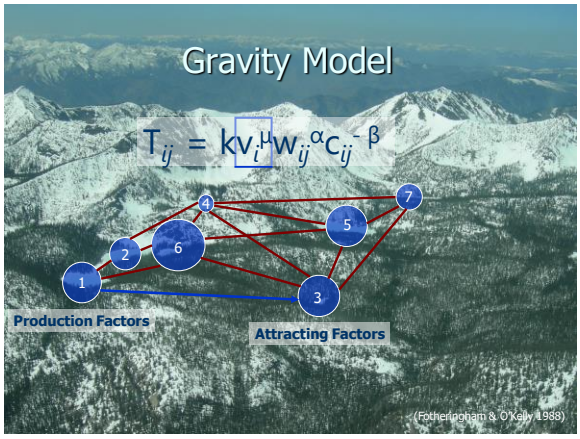
52



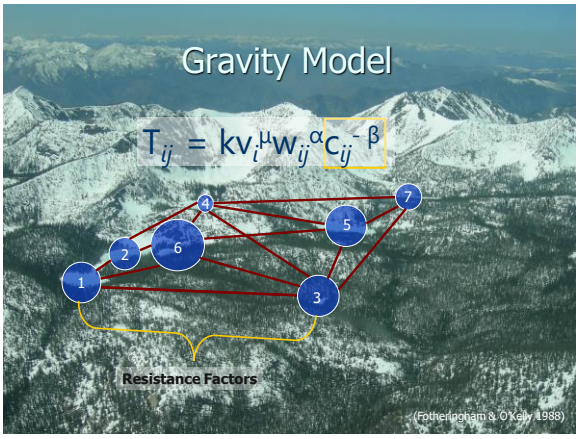
53



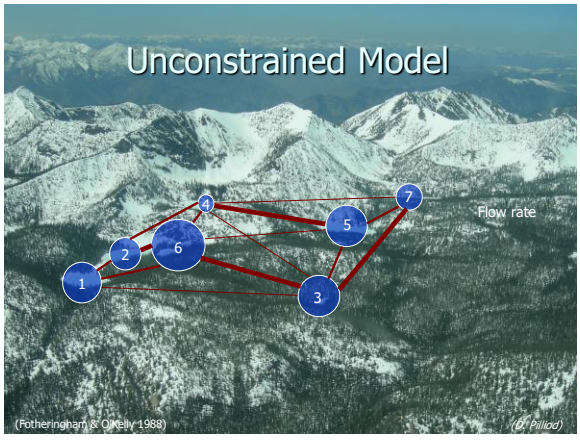
54



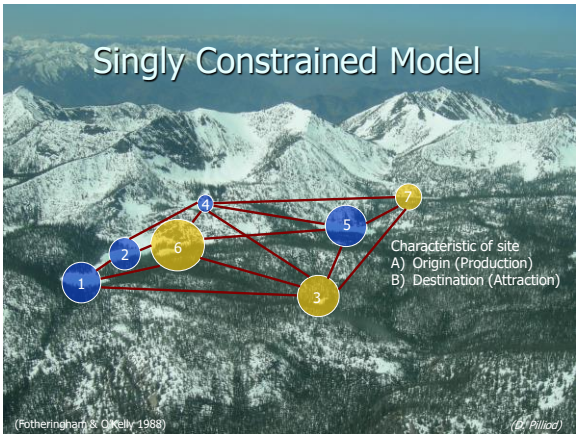
55



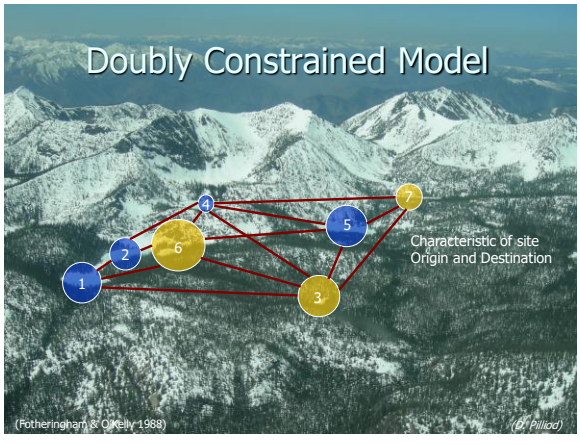
56



57



58



59

Trade-offs

Quality \ Quantity	Large	Small
High	Singly constrained	Doubly constrained
Low	unconstrained	

(Fotheringham & O'Kelly 1988)

60

"Calibration" of Gravity Equation

$$T_{ij} = kv_i^{\mu} w_{ij}^{\alpha} c_{ij}^{-\beta}$$

Unconstrained

$$\ln T_{ij} = \ln k + \ln(\mu v_i) + \ln(\alpha w_{ij}) - \ln(\beta c_{ij})$$

Linear regression – OLS
MLE

(Fotheringham & O'Kelly 1988)

61

"Calibration" of Gravity Equation

$$T_{ij} = kv_i^{\mu} w_{ij}^{\alpha} c_{ij}^{-\beta}$$

Singly Constrained

$$\ln T_{ij} = \ln k_i + (\ln(\mu v_i) + \ln(\alpha w_{ij}) - \ln(\beta c_{ij}))$$

Mixed Effects Models
MLE

(Fotheringham & O'Kelly 1988)

62

"Calibration" of Gravity Equation

$$T_{ij} = kv_i^{\mu} w_{ij}^{\alpha} c_{ij}^{-\beta}$$

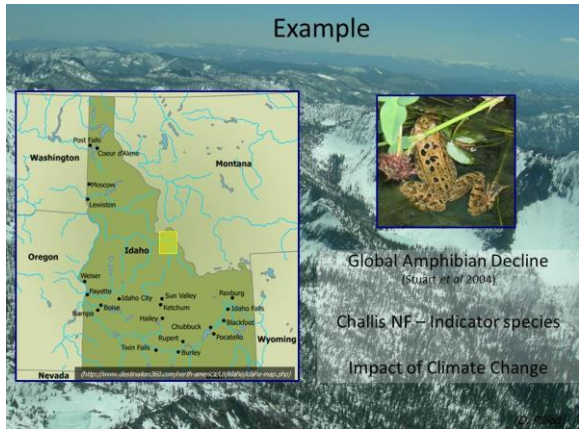
Doubly Constrained

$$\ln T_{ij} + \ln T_{ji} - \ln T_{jj} - \ln T_{ii} = \beta(\ln c_{ji} + \ln cji - \ln cii - \ln cjj)$$

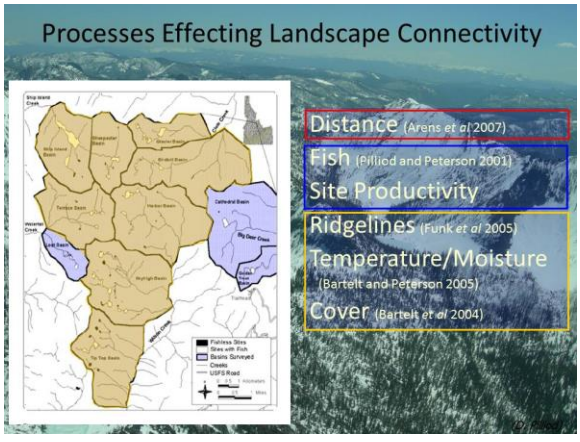
MLE

(Fotheringham & O'Kelly 1988)

63

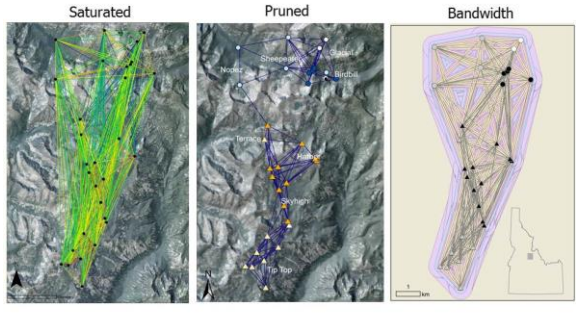


64



65

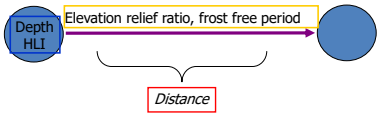
Validation



Significance test (1000 randomization); Sensitivity to missing sites (Jackknife)

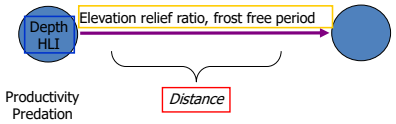
66

Connectivity is driven by landscape at site and between sites

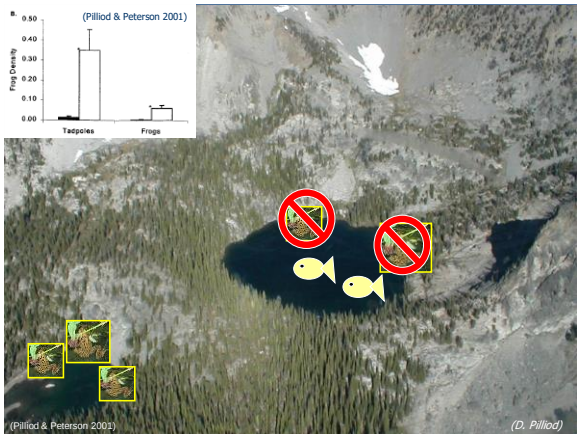


67

Connectivity is driven by landscape at site and between sites

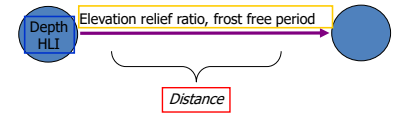


68



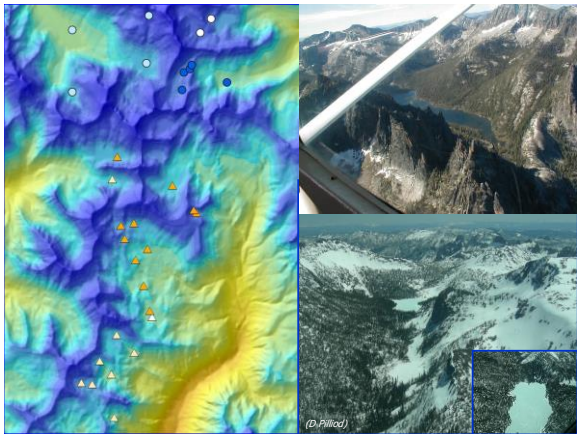
69

Connectivity is driven by landscape at site and between sites



Extremely short growing season
Ridgelines

70



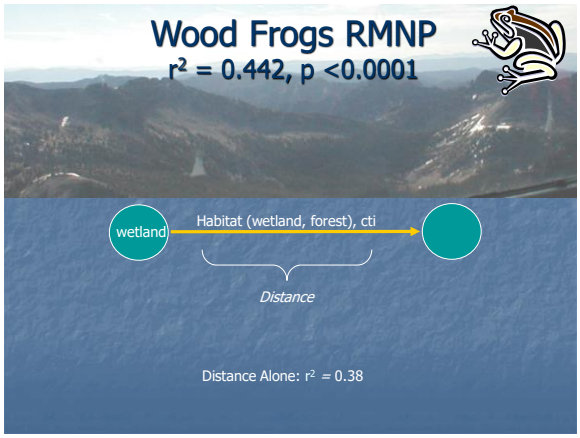
71

Wood Frog RMNP

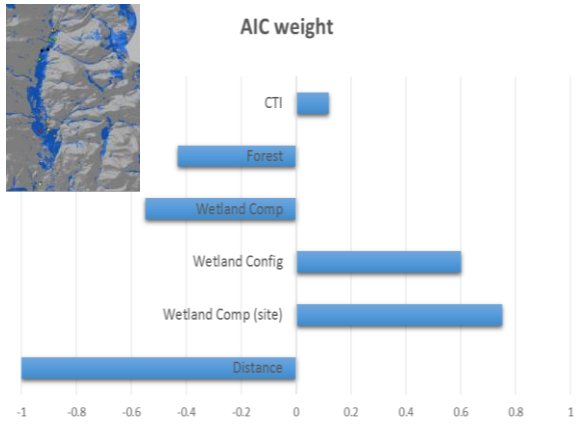
- High elevation
- Cold tolerant
- Breed in (beaver) ponds, vernal pools
- Impacted hydrology

Collaborators: Scherer, Muths, Oyler-M

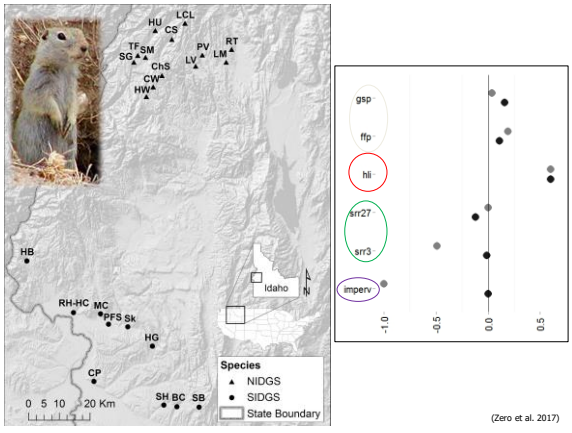
72



73

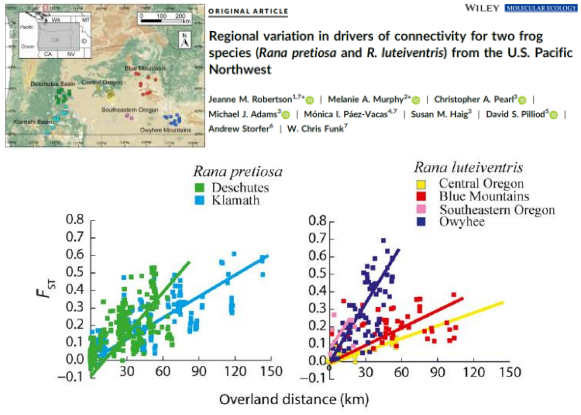


74

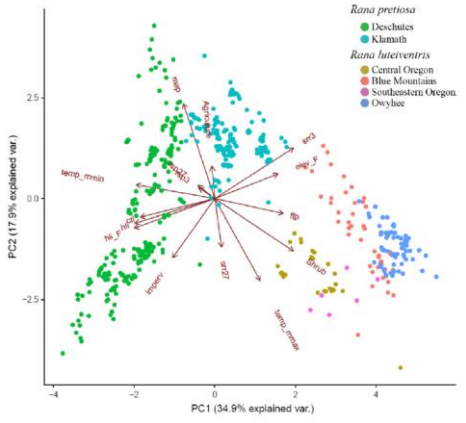


(Zero et al. 2017)

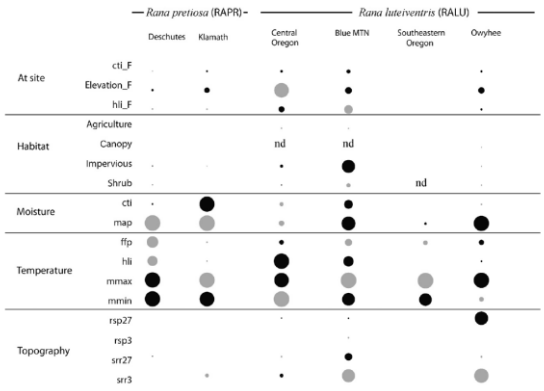
75



76

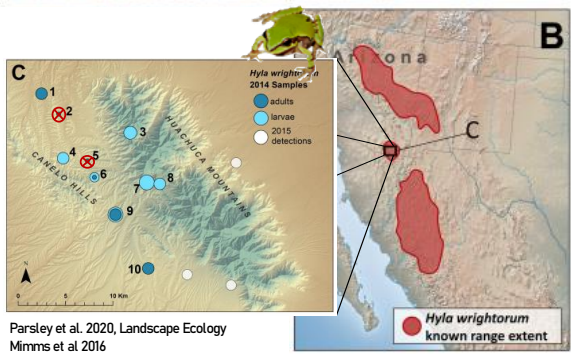


77



78

Local and landscape factors influence functional connectivity in an amphibian metapopulation

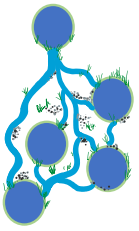


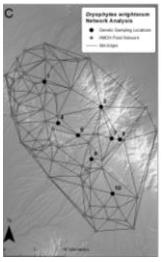
Parsley et al. 2020, Landscape Ecology
Mimms et al 2016

79

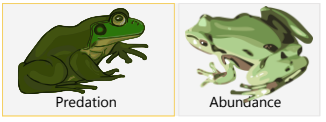
Primary productivity (SAVI)

Stream Distance

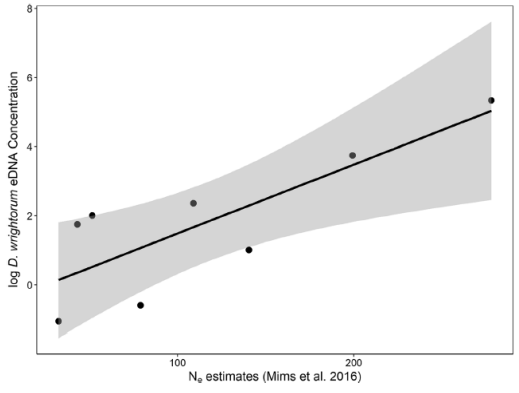




eDNA

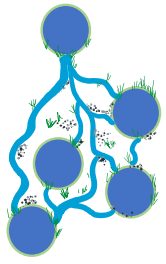


80



81

Model	K	$1\text{-}D_{PS}$ ΔAIC_C	AIC_C Weight	ΔBIC	BIC Weight
Streamdist	2	0.00	0.43	0.00	0.42
Predation	2	2.45	0.13	2.45	0.12
IBD	1	3.47	0.08	1.45	0.20
Hydroperiod	2	3.39	0.08	3.39	0.08
Stepping stone	2	3.81	0.06	3.81	0.06
Landscape moisture	2	4.59	0.04	4.59	0.04
Vegetation	3	5.18	0.03	7.20	0.01
Global	11	5.25	0.03	23.48	0.00
Connectedness	2	5.26	0.03	5.26	0.03
Abundance null	2	5.41	0.03	5.41	0.03
Water	4	5.60	0.03	9.65	0.00
Topography	3	6.90	0.01	8.93	0.00
Productivity	3	7.39	0.01	9.42	0.00



82

Spatial Autoregressive Model

- Probabilistic distribution
- No a priori weights or edge location
- Estimate resistance component parameters
- Spatially correlated residuals permitted
- Can account for missing values
- Prediction

(Peterson et al. 2019, Ecological Monographs)

83

Data format	Resistance covariates	A priori assumptions	Assumption-based outcomes	Resistance aggregation	Model
Contiguous 	**Node-based** Size, habitat quality, vegetation cover, distance between centroids	**First-order neighborhood structure** Rook, queen, percentage of boundary shared	**Resistance Covariates**	None	**Model-based** Estimates for individual resistance covariates
Non-contiguous 	**Node- or/and Edge-based** Habitat quality, vegetation cover, distance between node boundaries or centroids	**Edge location** Euclidean distance, buffered Euclidean distance	**Resistance Covariates**		
		Resistance values Expert opinion, literature review, empirical data, animal movement rates	**Edge location** Least-cost path, buffered least-cost path, multiple-least-cost path, Circuit-scape	Sum, difference, weighted product, mean, geometric mean, median of resistance values	**Model-based** Estimate for overall resistance values

84

b) Precision matrix
$\mathbf{Q} = (\mathbf{D}_W - \rho\mathbf{W}), \quad \rho = 0.90$

$\mathbf{Q}_{CAR}$ =

	V_1	V_2	V_3	V_4
V_1	8.0	-0.9	-2.7	-3.6
V_2	-0.9	1.0	0	0
V_3	-2.7	0	5.0	-1.8
V_4	-3.6	0	-1.8	6.0

c) Covariance matrix
$\boldsymbol{\Sigma} = \mathbf{Q}^{-1}$

$\boldsymbol{\Sigma}_{CAR}$ =

	V_1	V_2	V_3	V_4
V_1	0.55	0.49	0.46	0.47
V_2	0.49	1.44	0.42	0.42
V_3	0.46	0.42	0.62	0.46
V_4	0.47	0.42	0.46	0.59

d) Partial correlation matrix
$\mathbf{k}_{i,j} | . = \text{corr}(\varepsilon_{i|.}, \varepsilon_{j|.}) = -\mathbf{Q}_{i,j} / \sqrt{\mathbf{Q}_{i,i}\mathbf{Q}_{j,j}}$

$\mathbf{k}_{CAR}$ =

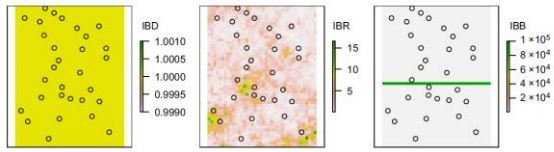
	V_1	V_2	V_3	V_4
V_1	1	0.32	0.43	0.52
V_2	0.32	1	0	0
V_3	0.43	0	1	0.33
V_4	0.52	0	0.33	1

e) Correlation matrix
$\text{Corr}_{i,j} = \boldsymbol{\Sigma}_{i,j} / \sqrt{\boldsymbol{\Sigma}_{i,i}\boldsymbol{\Sigma}_{j,j}}$

Corr_{CAR} =

	V_1	V_2	V_3	V_4
V_1	1	0.55	0.80	0.83
V_2	0.55	1	0.44	0.46
V_3	0.80	0.44	1	0.77
V_4	0.83	0.46	0.77	1

85



86

Recommendations

- Clearly ID research questions
- Choose an adequate study design
- Sensitivity test (missing nodes, band width)

(Murphy *et al* 2015; *Landscape Genetics: Concepts, Methods, Applications*)

87

Graphs in Landscape Genetics

- Pairwise geneflow data inherently graph-like in structure
- Inferences from graph structure
- Topological comparison (hypothesis testing)
- Modeling flow through a network

88

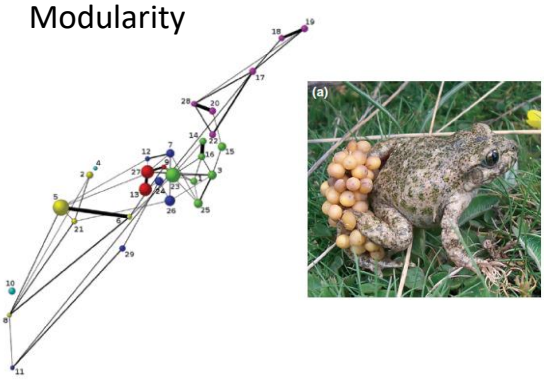
Lab – Columbia Spotted Frog

- Part 1 - Conceptual overview of graph metrics.
- Part 2 – Gravity model analysis
- Part 3 – Gravity models and model selection with your own models

89

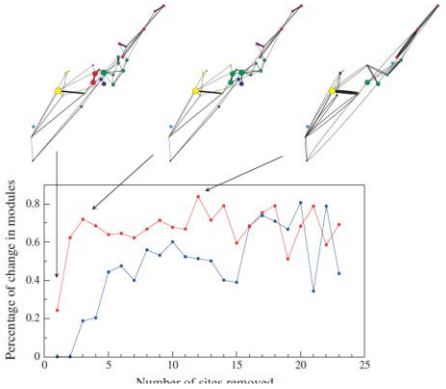
90

Modularity



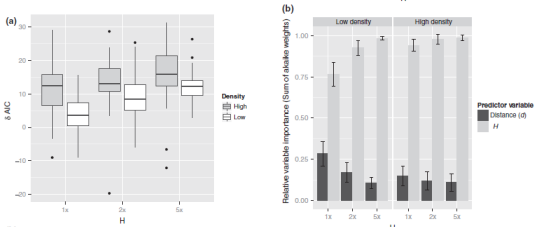
(Albert *et al.* 2014, Ecology Letters)

91



92

Gravity of Pollination



(Dileo *et al.* 2014; Molecular Ecology)

93

Are Graphs Spatial?

Aspatial

Spatial

Classic Graph Theory – abstract

Intermediate – direction, flow

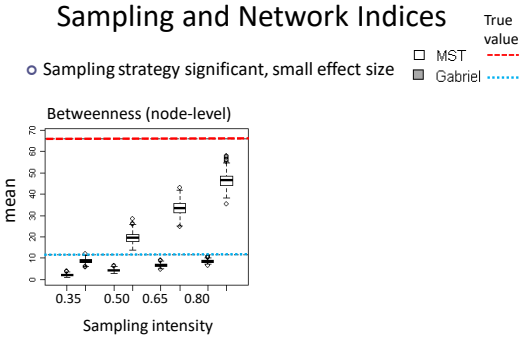
Extreme Case

Objects – location, size, shape, composition

Relationships – location, size, shape, width, length

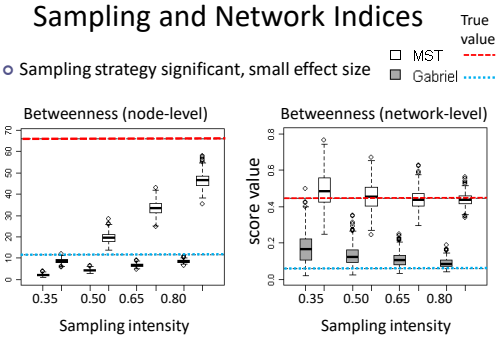
94

Graph Structure: Sampling and Network Indices



95

Graph Structure: Sampling and Network Indices



96

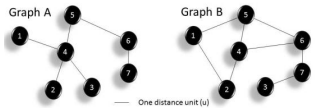
Local Metrics

By Node

Node	Degree		Betweeness		Closeness	
	A	B	A	B	A	B
1	1	2	0	0.5	0.071	0.077
2	1	2	0	0.5	0.071	0.077
3	1	1	0	0	0.071	0.059
4	4	3	12	3.5	0.111	0.100
5	2	3	8	3.5	0.100	0.100
6	2	3	5	8	0.077	0.111
7	1	2	0	6	0.056	0.083

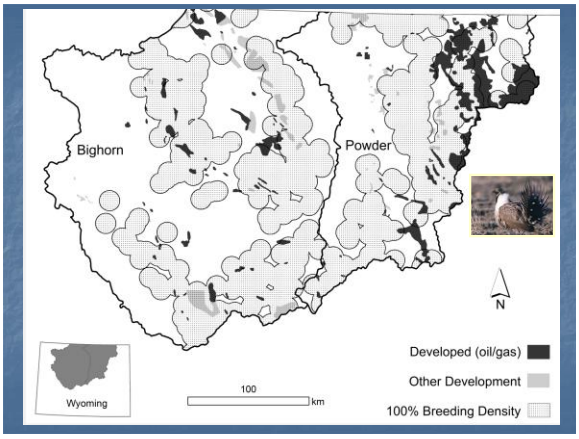
Summarized by Graph

Metric	A	B
Freeman's Degree Centrality	0.53	0.17
Mean Closeness	0.08	0.09
Mean Betweeness	3.57	3.14

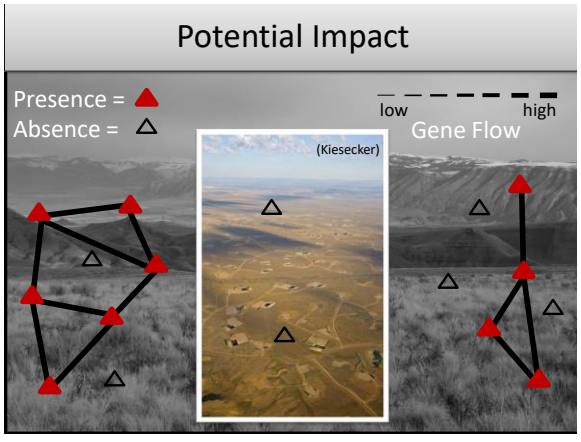


(Murphy *et al.* 2015)

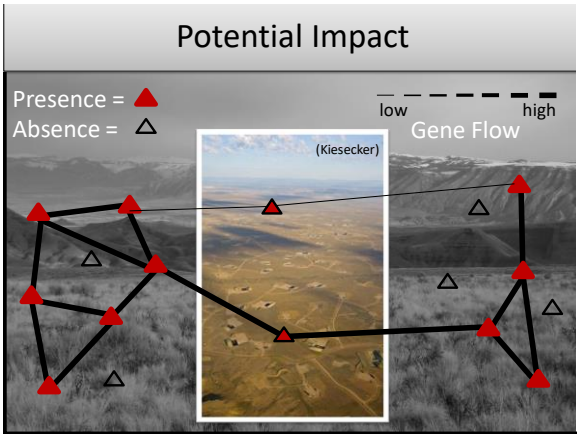
97



98



99



100