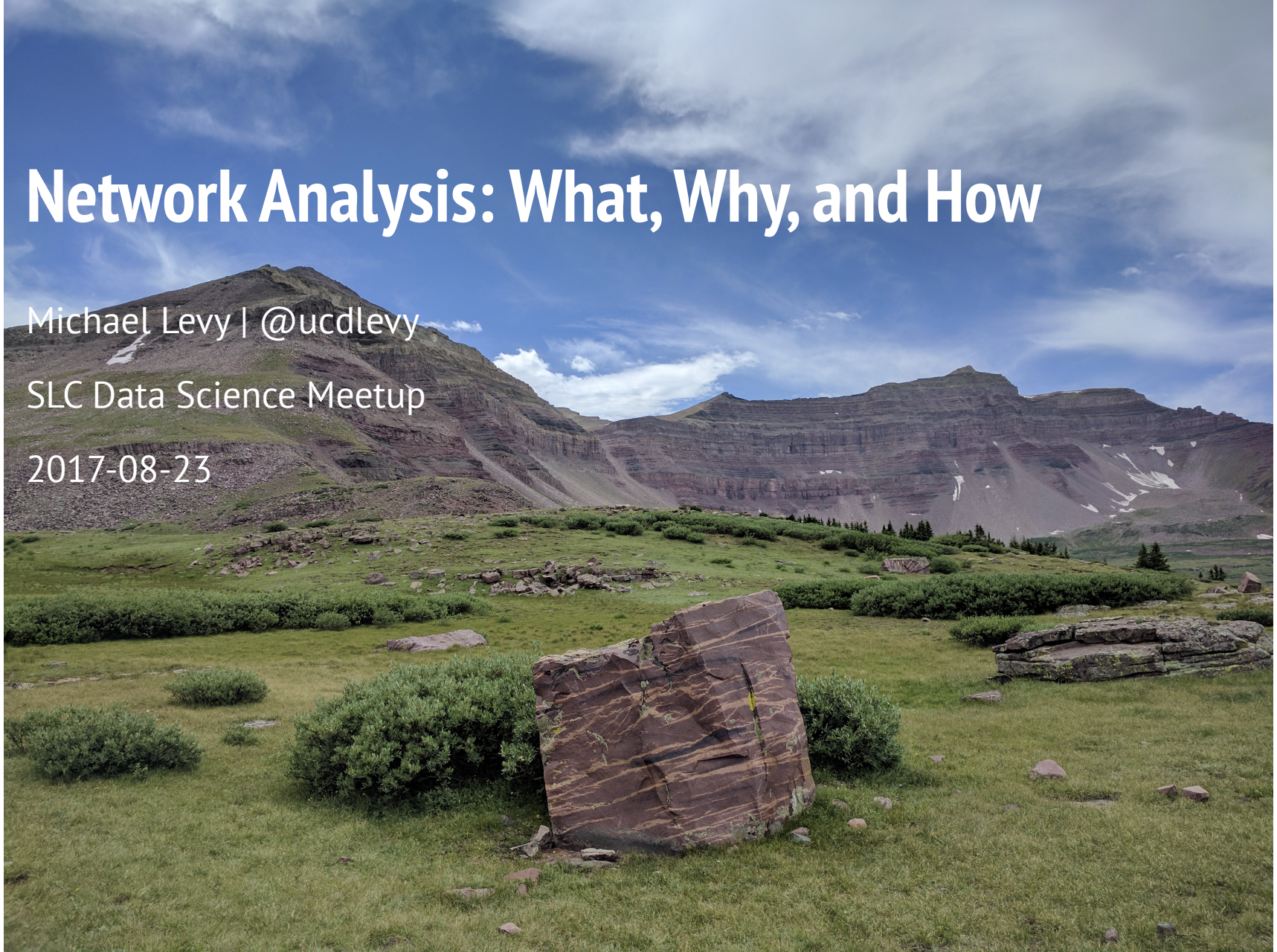
Network Analysis: What, Why, and How

Michael Levy | @ucdlevy

SLC Data Science Meetup

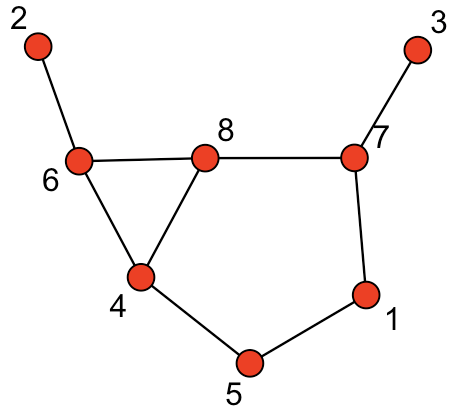
2017-08-23



Agenda

1. What is a network?
2. Real-world examples
3. Network statistics
4. Four case studies
    1. Descriptive statistics
    2. Conditional uniform random graph tests
    3. Network autocorrelation models
    4. Exponential random graph models
5. Tools

What makes a network?



Points	Lines	Tradition
nodes	links	CS
vertices	edges	math
actors	ties	social science

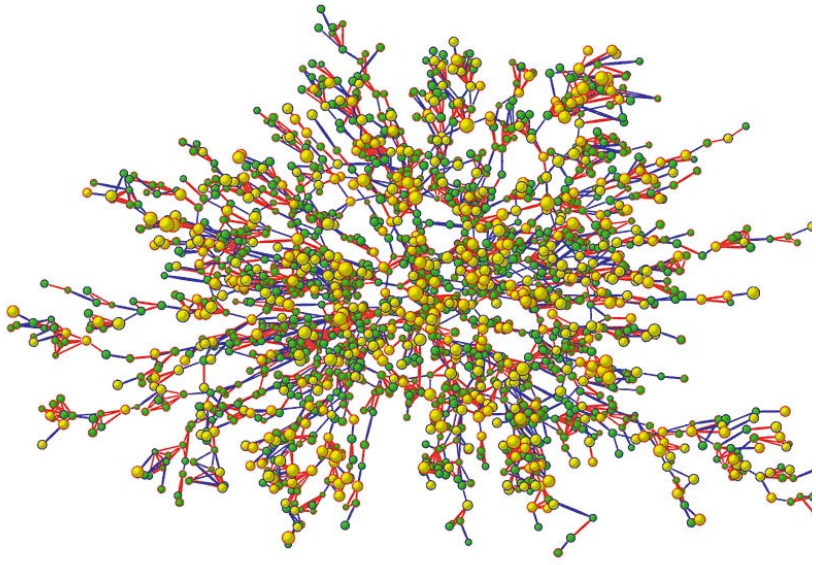
Example networks

Nodes	Edges
People	Friendship / Number of emails / Virus transmission
Computers	Data transfer / Compatibility
Cities	Highways / Migration volume / Gov't collaboration
Proteins	Interact / Share sequence
Functions	Calls

Real-world examples


Facebook Friendships
facebook
December 2010

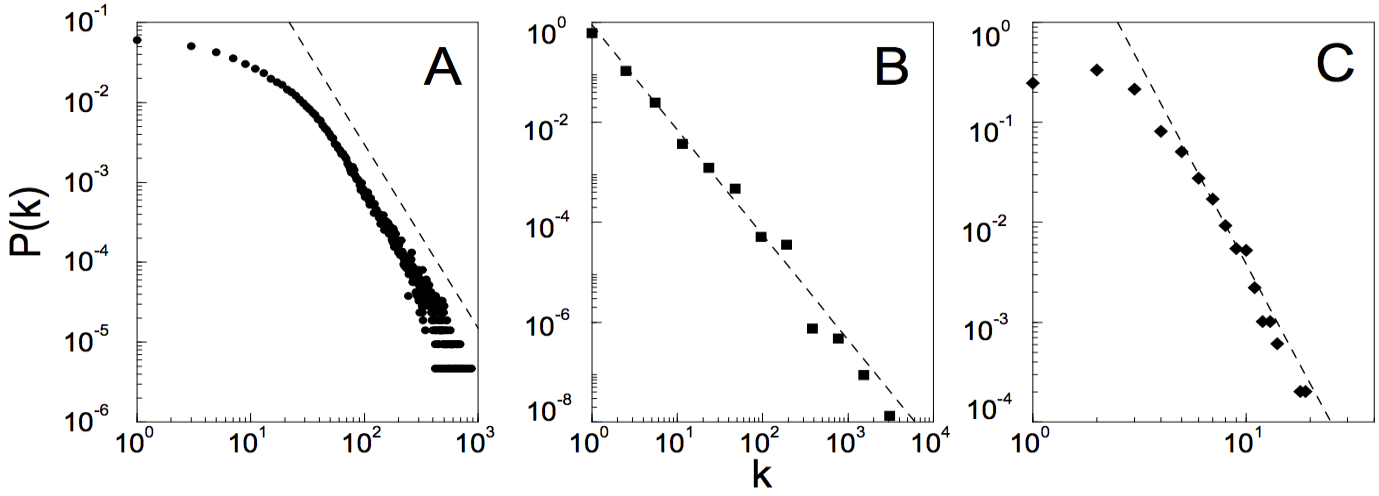
Contagion of Obsesity



- Nodes = People
  - Yellow = obese
  - Green = non-obese
- Edges = Relationships
  - Purple = friendship or marriage
  - Orange = related

Christakis & Fowler 2007

Scale Invariance

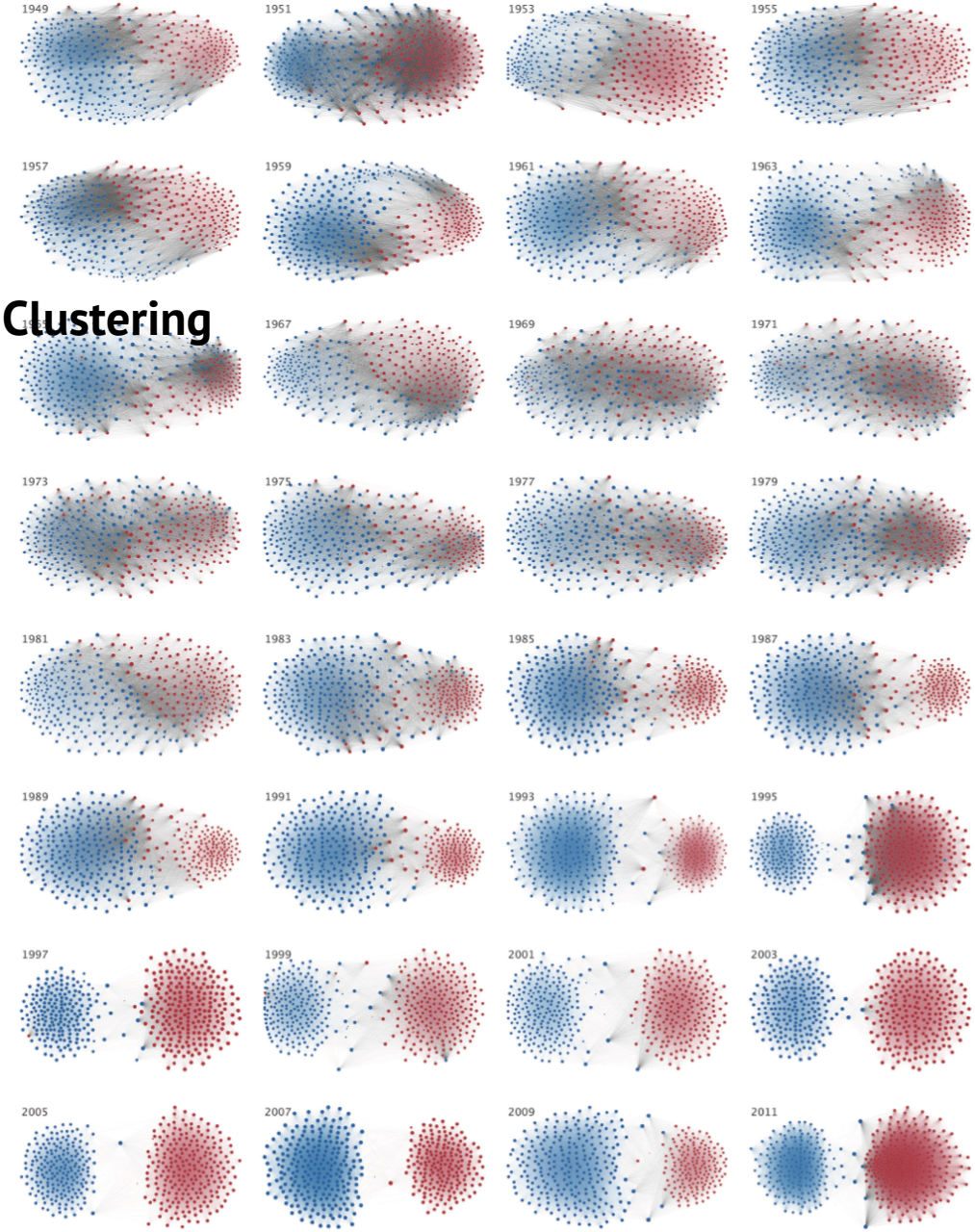


A = actor collaboration, B = www links, C = power stations

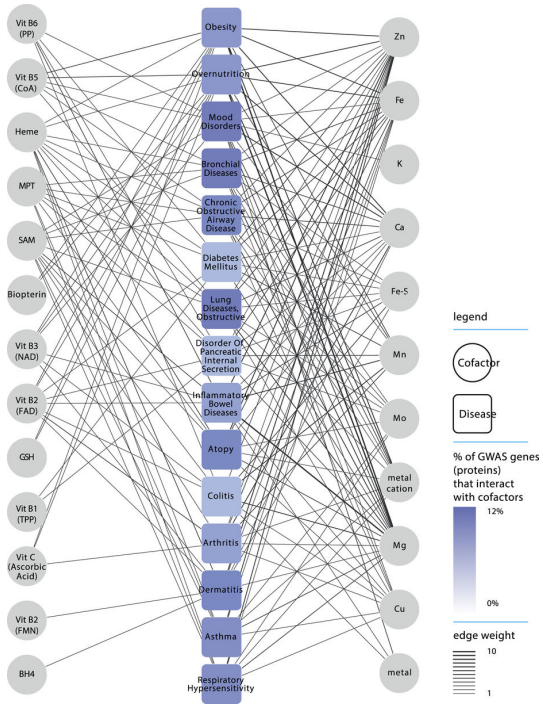
Barabasi & Albert 1999

Congressional Clustering

Andris *et al.* 2015

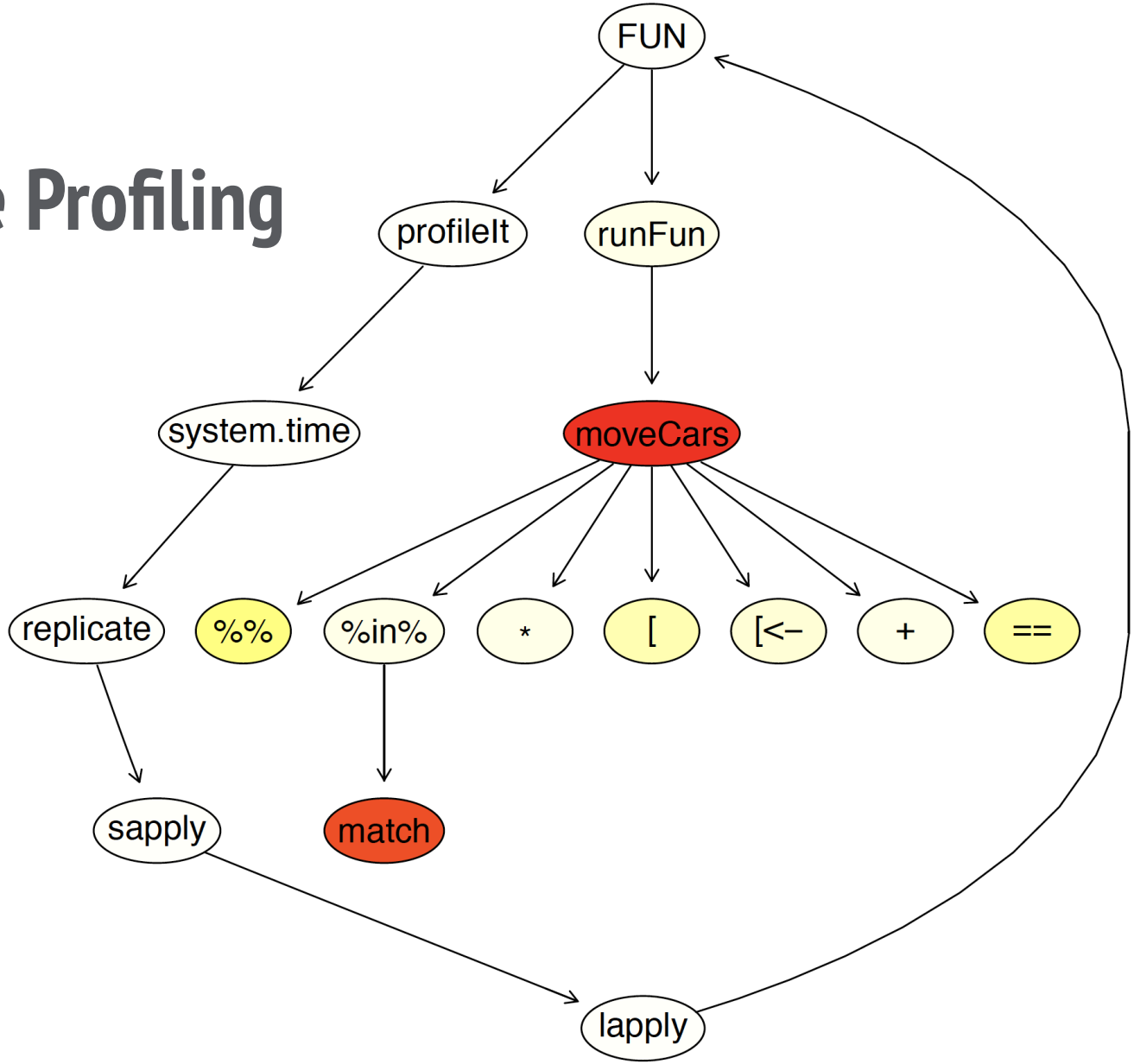


Protein-Cofactor-Disease Interactions



Scott-Boyer *et al.* 2016

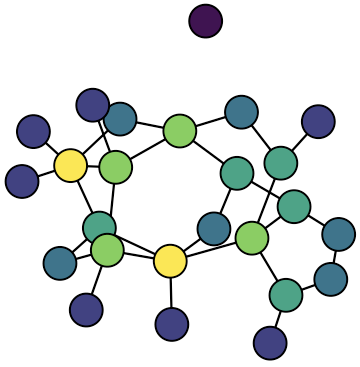
Code Profiling



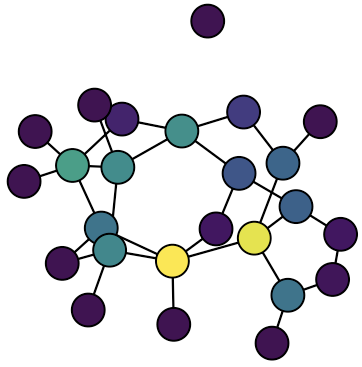
Individual level statistics

Centrality

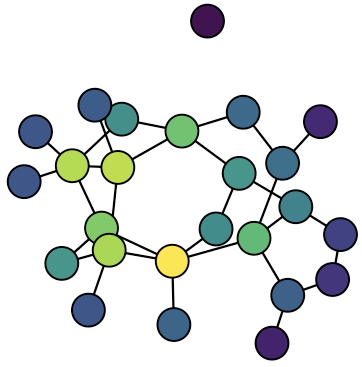
Degree



Betweenness



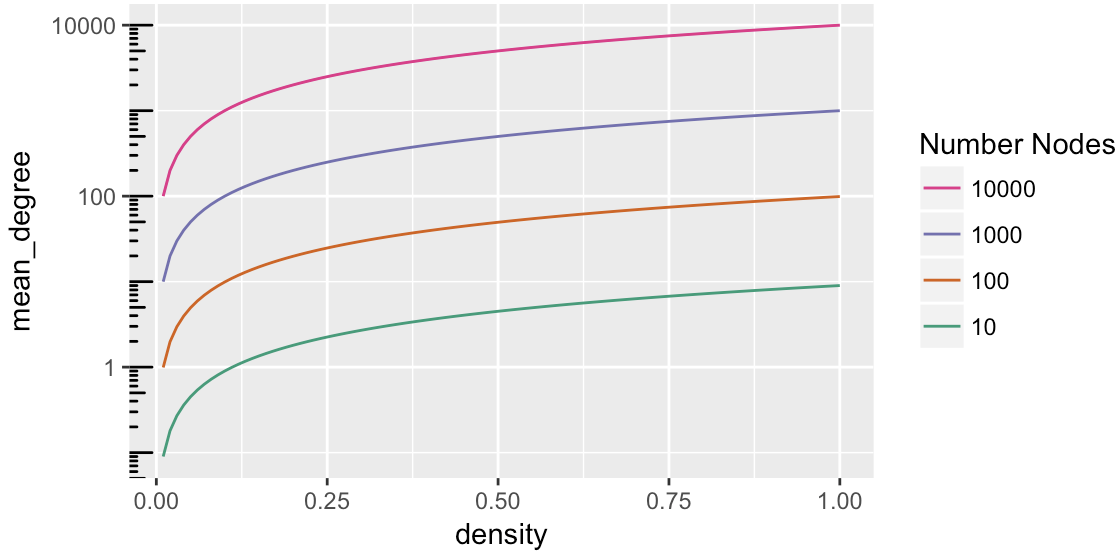
Eigenvector



Network level statistics

Density

- Density = Fraction of possible edges present, $\in [0, 1]$
- Mean Degree = Average degree of nodes, $\in [0, N]$

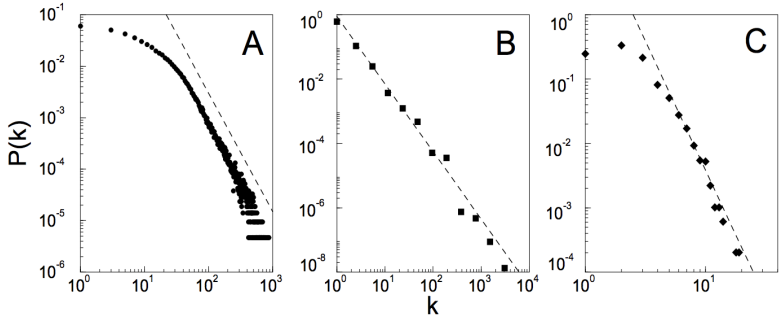


Centralization

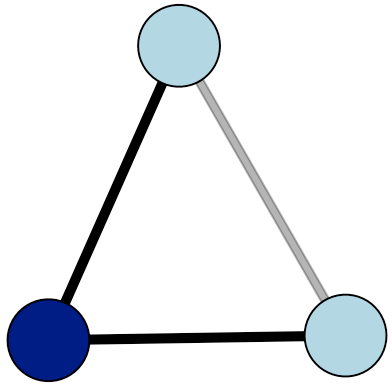
For centrality measure C on vertex v:

$$Centralization = \sum_{v} max(C(v)) - C(v)$$

Variance of centrality also reasonable

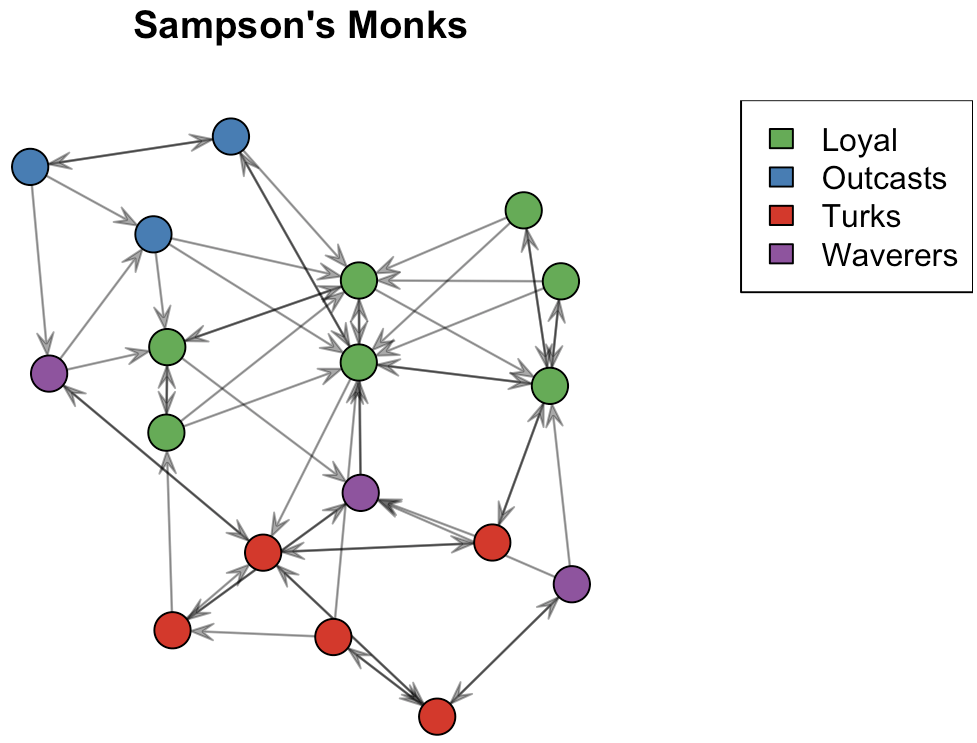


Clustering coefficient



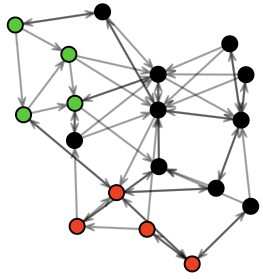
- $\frac{3 \times triangles}{twopaths}$
- Measure of how tightly bound neighborhoods are.

Homophily

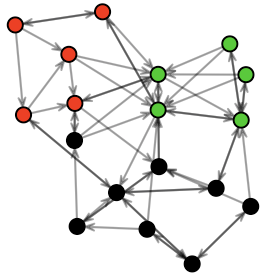


Modularity

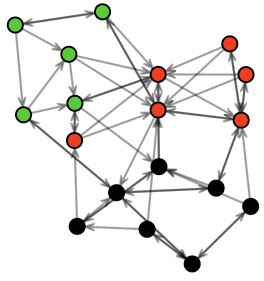
Edge Betweenness: 0.235



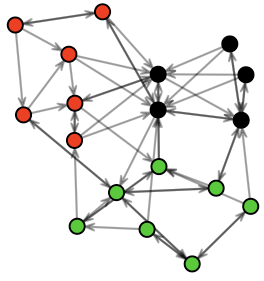
Fast Greedy: 0.264



Walk Trap: 0.284

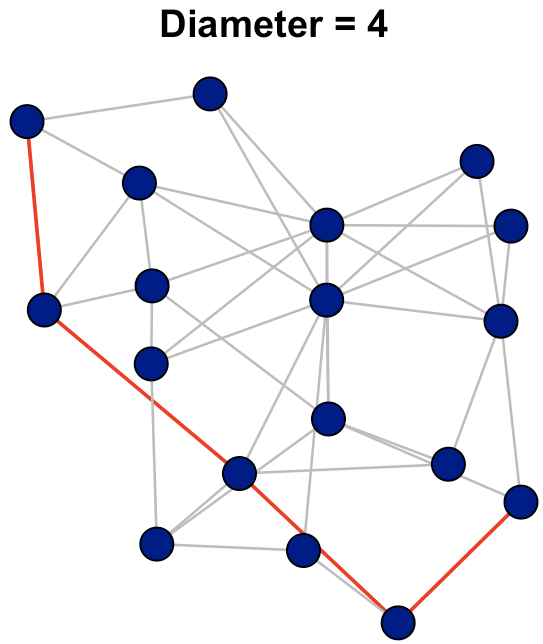


Spin Glass: 0.273



Diameter

Longest-shortest path between two nodes



More Measures

- Average path length
    - Harmonic mean handles disconnected nodes
- Connectedness
- Hierarchy
- Efficiency
- Least-upperboundedness

Variations

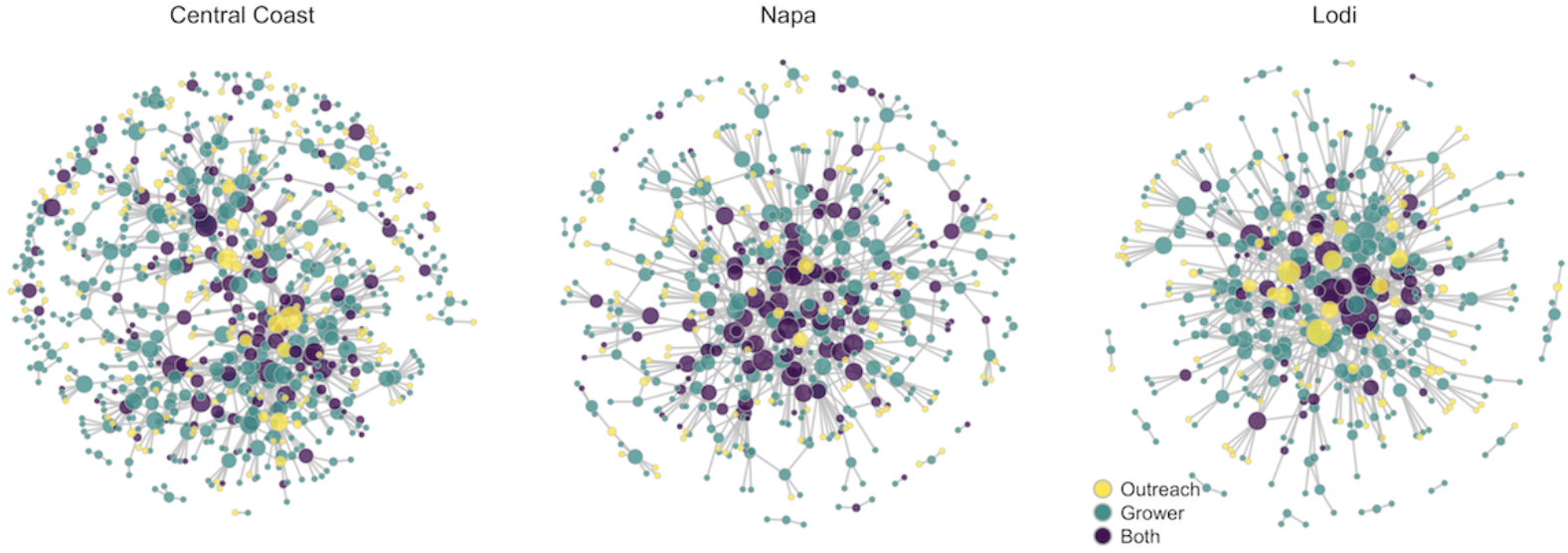
- Special Cases
  - Trees
  - Bipartite networks
- Generalizations
  - Valued/count edges
  - Directed edges
  - Dynamic networks
  - Dynamics on networks

Case Studies

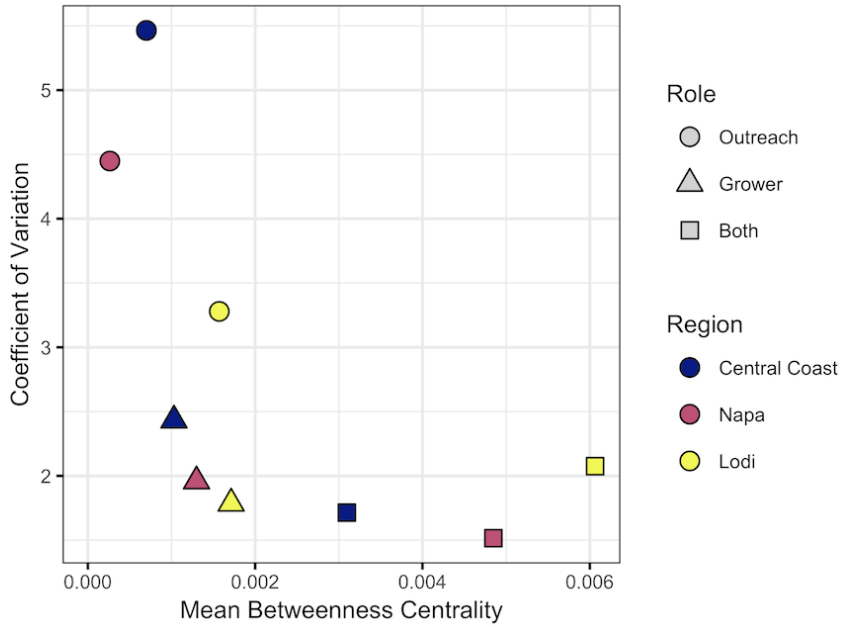
1. Viticulture management: Descriptive statistics
2. Mental models of sustainable agriculture: Conditional uniform random graph (CUG) tests
3. Pre-teen obesity: Network autocorrelation models (LNAM)
4. Mountain lion prey sharing: Exponential random graph models (ERGM)

Viticulture management

Three regional networks, three professional categories

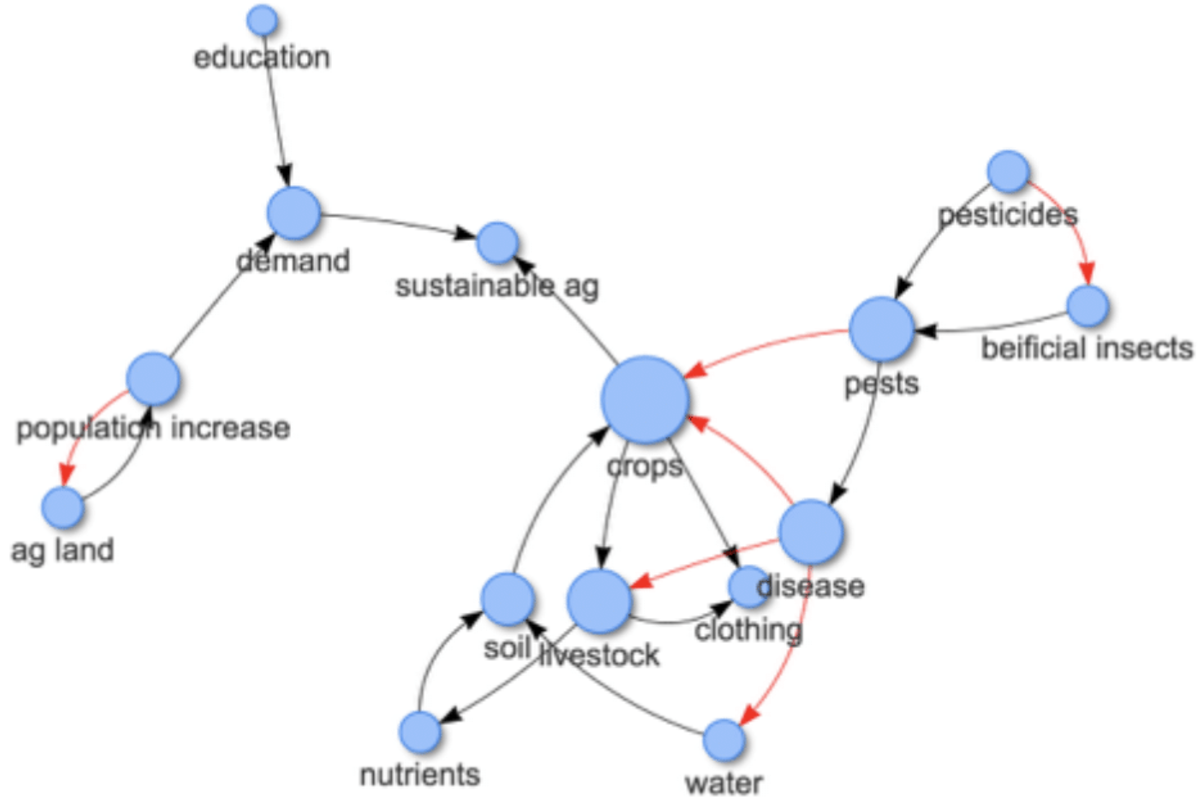


Outreach professionals who grow control information flow



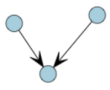
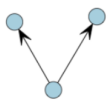
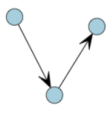
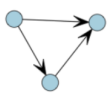
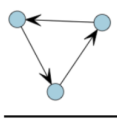
Mental models of sustainable agriculture

Network of perceived causal relationships

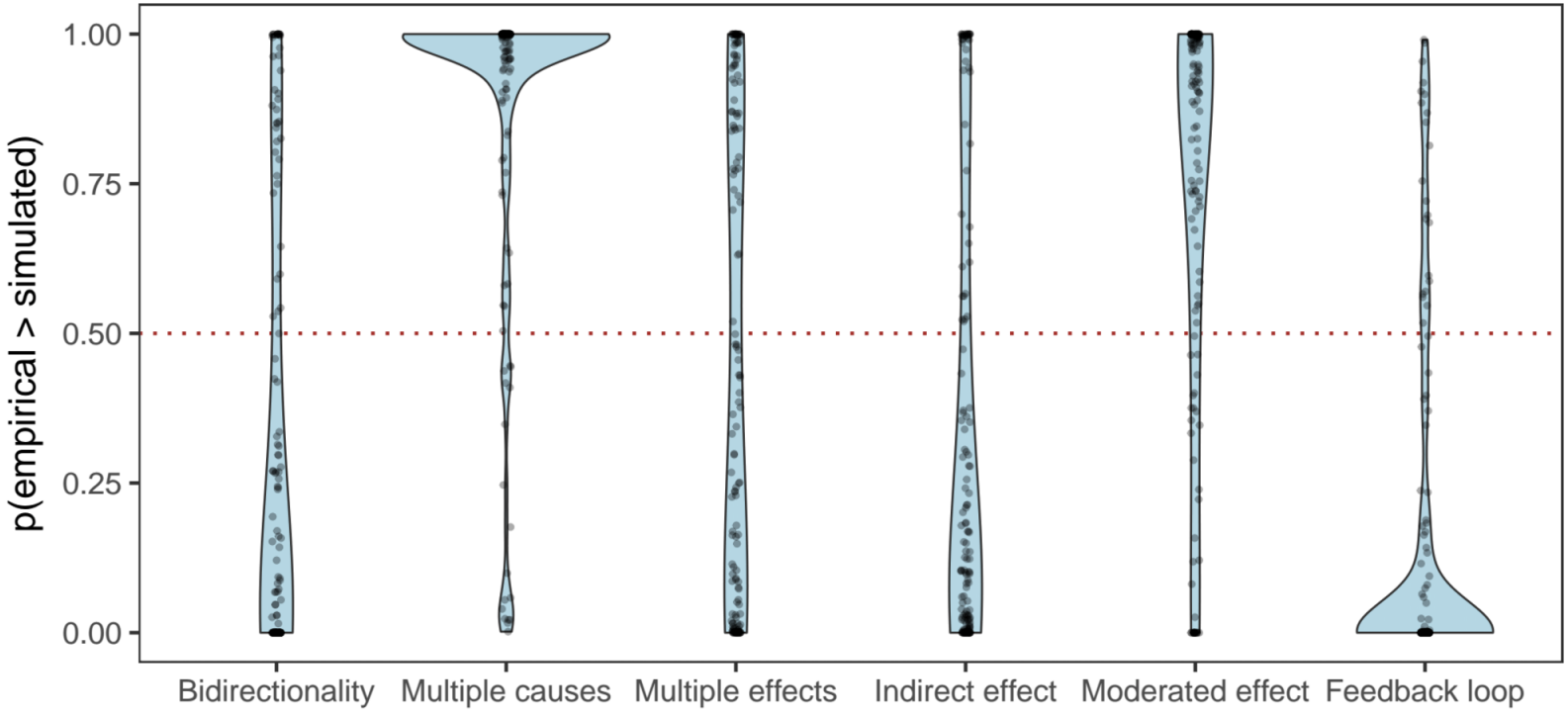


Causal patterns as network motifs

Table 1: Six motifs that form the building blocks of networks and fundamental patterns of causality.

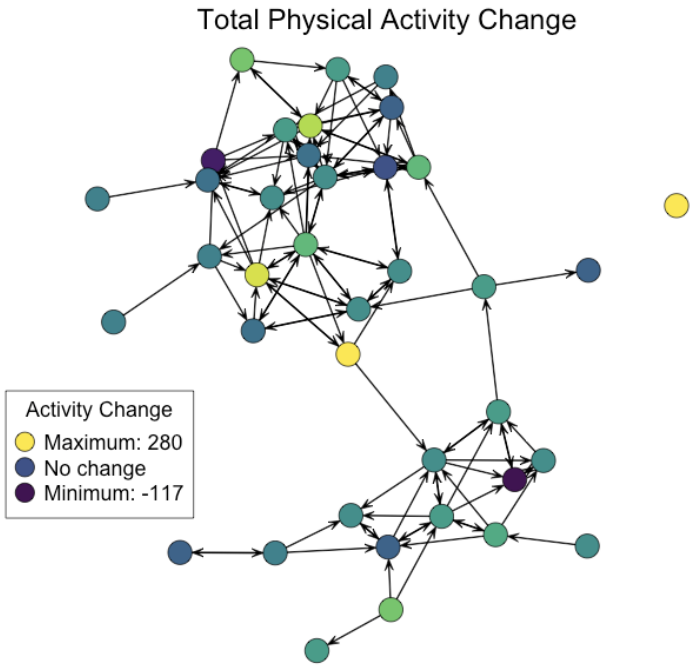
Motif	Causality	Network Structure	Cognitive Map Metric	Hypothesis	Emprical Prevalence
	Bidirectionality	Reciprocal Pair	Hierarchy (−)	Rare	Rare
	Multiple causes	In Star	Transmitter Variables	Common	Very common
	Multiple effects	Out Star	Receiver Variables	Rare	Balanced
	Indirect effect	Two Path	Ordinary Variables	Rare	Rare
	Moderated effect	Transitive Triple	Hierarchy	?	Common
	Feedback loop	Cyclic Triple	Hierarchy (−)	Rare	Very rare

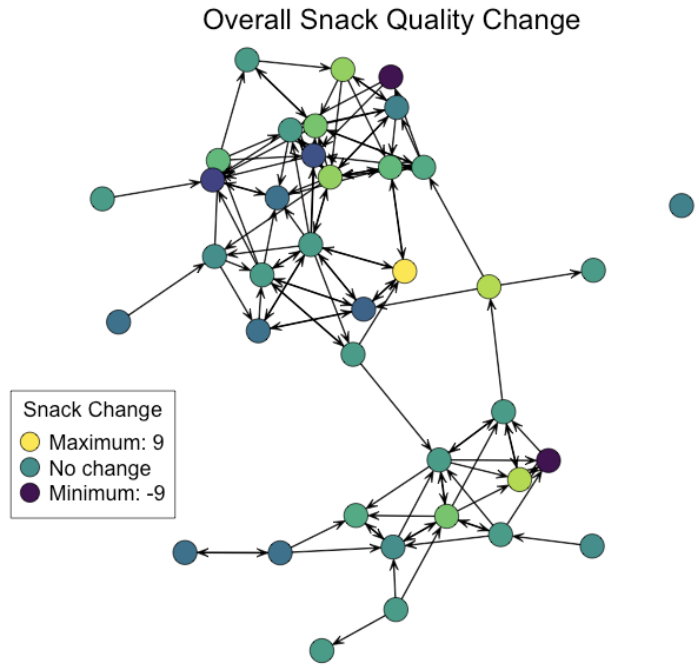
Representation of causal patterns vs. chance



Pre-teen obesity

10-12 y.o. girls at obesity-intervention camp





Evidence for social support but not contagion

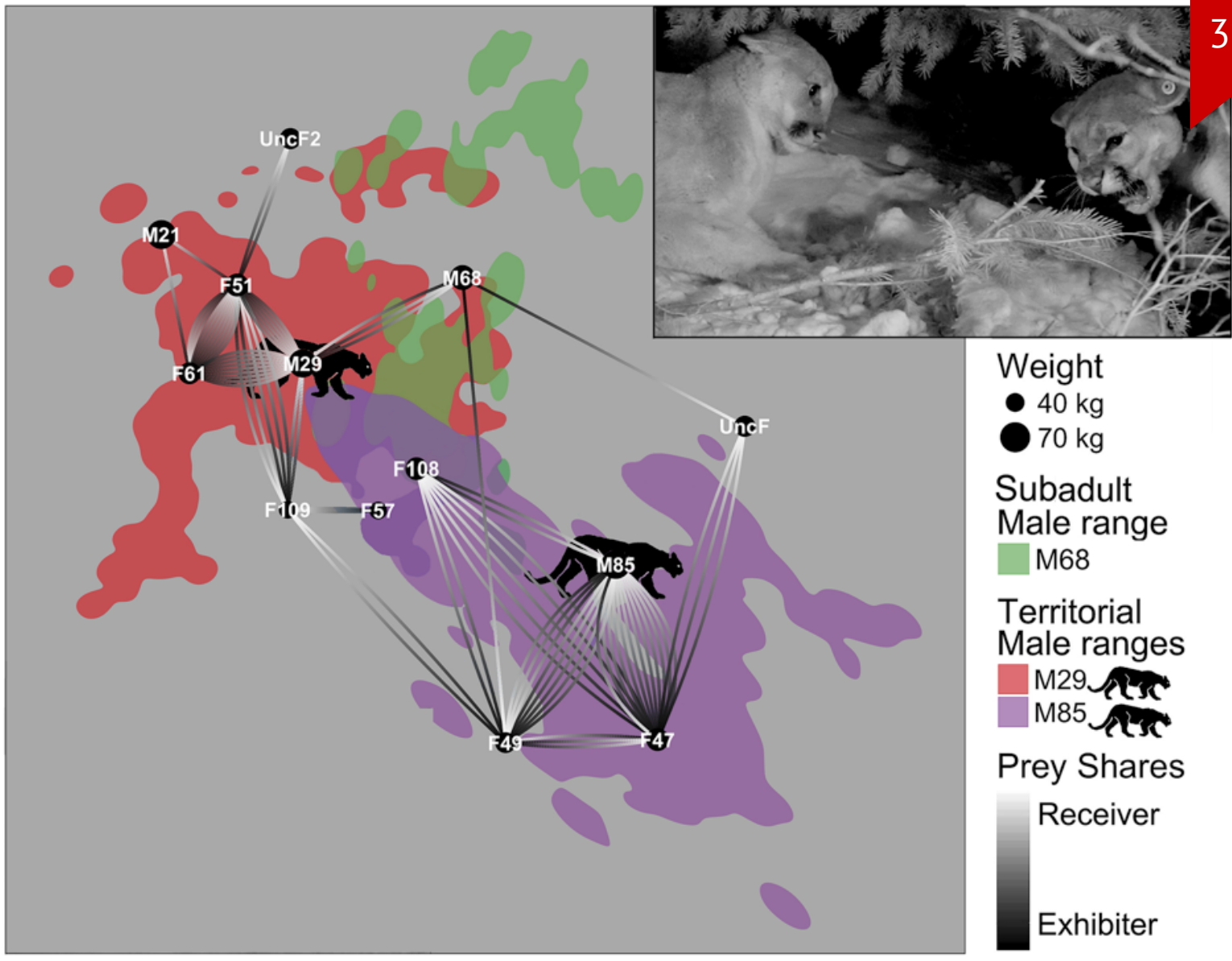
LNAM of snack quality change

```
        Estimate Std. Error Z value Pr(>|z|)
age     -0.08404    0.08233  -1.021   0.3074
friends  0.46449    0.26427   1.758   0.0788 .
rho1.1  -0.04893    0.23314  -0.210   0.8338
```

LNAM of physical activity change

```
        Estimate Std. Error Z value Pr(>|z|)
age       5.8990     2.0673   2.853  0.00432 **
friends  10.4995     6.2888   1.670  0.09501 .
rho1.1   -0.2564     0.2113  -1.213  0.22511
```

Mountain lion prey sharing


UncF2
M21
F51
M68
F61
M29
UncF
F108
F109
F57
M85
F49
F47
Weight
40 kg
70 kg
Subadult
Male range
M68
Territorial
Male ranges
M29
M85
Prey Shares
Receiver
Exhibiter

Statistical evidence of social structure

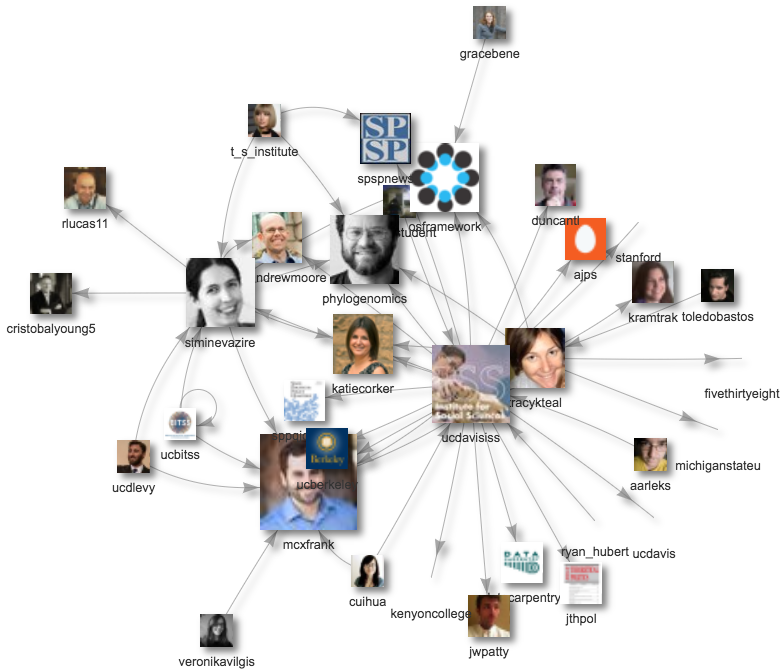
	Binary edges	Valued edges
Land-tenure and kinship hypotheses		
Relatedness	2.22 (2.09)	0.25 (0.68)
Spatial Overlap	1.44 (1.09)	0.88* (0.37)
Male Receiving	1.02 (0.65)	0.42** (0.16)
Male Sharing	-0.67 (0.47)	-0.68** (0.26)
Social behavior hypotheses		
Direct reciprocity	2.04* (0.83)	0.78* (0.31)
Hierarchical reciprocity (transitivity)	0.91* (0.46)	0.28* (0.14)
Generalized reciprocity (cyclicality)	-0.18 (0.36)	-0.09 (0.14)
Basic Network Attributes		
Density	-4.02*** (0.91)	0.48* (0.22)
Any Shares		-4.42*** (0.42)
Even-Distribution Receiving	2.59. (1.52)	
Even-Distribution Sharing	-0.45 (0.97)	

Tools

R

- igraph
- statnet (network, sna, ergm, networkDynamic): Statistical analysis of networks (ERGM, CUG, lnam, etc.)
- RSiena: Statistical models on temporal data
- visNetwork: Dynamic visualization from R via D3

Select by id


gracebene
t_s_institute
spspnews
rlucas11
orgframework
duncantl
andrewmoore
phylogenomics
ajps
stanford
kramtrak
toledobastos
cristobalyoung5
sminevazire
katiecorker
racykteal
fivethirtyeight
ucdavisiss
ucbitss
michiganstateu
ucdlevy
aarleks
mcxframk
ryan_hubert
ucdavis
cuihua
carpentry
kenyoncollege
jthpol
jwpatty
veronikavilgis

Python

- networkX
- igraph
- ???

GUI

- Gephi: Open-source visualization and analysis

- Pajek
- ORA

More

- For a comprehensive list of tools, see the curated Awesome Network Analysis list: https://github.com/briatte/awesome-network-analysis
- These slides, including some R code: https://github.com/michaellevy/slc_datasci_talk
- Feel free to get in touch with me:
  - @ucdlevy
  - michael.levy@healthcatalyst.com